

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR



TESIS DOCTORAL

**Estudio por RMN del reconocimiento molecular carbohidrato-proteína con
discriminación diferencial de anómeros en equilibrio**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

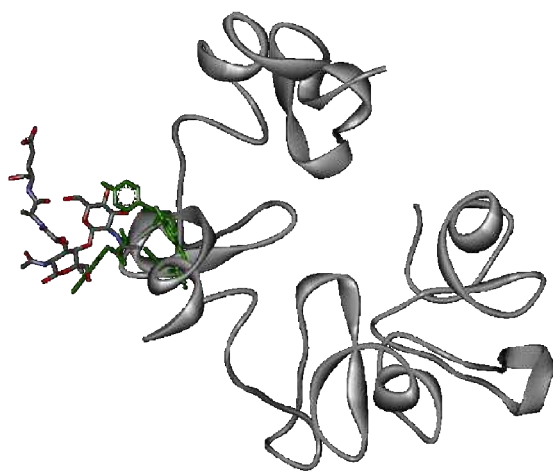
Khouzaima El Biari

Directores

Javier Cañada Vicinay
Jesús Jiménez Barbero

Madrid, 2014

**Estudio por RMN del reconocimiento molecular
carbohidrato/proteína con discriminación
diferencial de anómeros en equilibrio**



TESIS DOCTORAL

Khouzaima El Biari

Madrid 2014

UNIVERSIDAD COMPLUTENSE DE MADRID



Facultad de Ciencias Biológicas

Departamento de Bioquímica y Biología Molecular



**Estudio por RMN del reconocimiento molecular
carbohidrato/proteína con discriminación
diferencial de anómeros en equilibrio**

Memoria para optar el grado de doctor presentada por

Khouzaima El Biari

Madrid, 2008

Esta Tesis ha sido realizada en el Centro de Investigaciones Biológicas
(CSIC)

Bajo la codirección de

Dr. Javier Cañada Vicinay

Dr. Jesús Jiménez Barbero

A mis padres y a mis hermanos

No sé si alguien puede llegar a imaginarse la importancia sentimental y la profunda liberación que tiene acabar y presentar, por fin, esta Tesis Doctoral. Pero no sólo para mí, sino también para los míos, sobre todo para quienes me han alentado tanto a lo largo de este largo viaje, animándome sin pausa en esta etapa de mi vida. Es que han pasado unos seis años desde que llegue a España, precisamente al Centro de Investigaciones Biológicas....Durante todo este tiempo, han sido muchas las personas que han ayudado de una manera u otra en la elaboración de este trabajo. Así que espero ser capaz de nombrarlos a todos y, sobre todo, transmitir mi más sincero y profundo agradecimiento.

En primer lugar, me gustaría agradecer a mis directores de tesis, Javier y Jesús, el haberme brindado la oportunidad de realizar esta Tesis y de conocer el mundo de la Ciencia. A Javier, me gustaría darle las gracias de un modo especial por su constante entrega y dedicación. Sin tu ayuda y constantes ánimos no hubiera podido acabar este trabajo. Gracias también por la paciencia mostrada en nuestras interminables ‘discusiones’ científicas, los conocimientos aportados y por no rendirte nunca, incluso en los meses difíciles cuando se terminó la beca y me tocó estar en Marruecos, rescatándome telefónicamente para seguir con el trabajo cuando todo se tornaba de un color muy negro. Y a ti Jesús, gracias por haber confiado en mí y por la ayuda y consejos proporcionados durante la elaboración de esta Tesis, te agradezco especialmente que hayas considerado siempre mi opinión y el haberme buscado y proporcionado el contrato que me ha ayudado a llevar este proyecto a término.

A Mourad me gustaría agradecerle muchísimas cosas, pero ni siquiera sé por donde empezar. Gracias por tu apoyo permanente, la comprensión de mis locuras, los ánimos constantes, la múltiple ayuda y conocimientos científicos aportados, por ayudarme a superar momentos difíciles...En suma, gracias por tu gran amistad.

Ha sido un gran placer compartir profesionalmente estos años con toda la gente con la que ha compartido mi paso por el laboratorio. De mis primeros años, me gustaría agradecer a Iñaki, Luis, María y Carmen su ayuda en la adaptación tanto al laboratorio como a la inmensa ciudad de Madrid, su sentido del humor y las largas conversaciones que mantuvimos me abrieron los ojos a otro mundo. A Ángeles, Dolores, Ana, Victoria Joao, Lidia y Filipa gracias por las cientos de cafés matutinos, las charlas, las fiestas, las risas que nos hemos echado, vuestra forma de ser tan ‘auténtica’. También querría expresar mi agradecimiento a Álvaro Pilar, Blanquita, Lucas, Anita, Silvia y Beatriz por

su maravillosa forma de ser, por todo lo compartido durante años, por su trato humano, comprensión y apoyo no sólo en lo profesional, sino también en lo personal.

A Dra Eva Batanero, mi tutora durante la tesis, por estar siempre dispuesta a ayudarme y aconsejarme aunque desapareciese durante años. Al departamento de Bioquímica y Biología Molecular de la UCM. Al personal del CIB, especialmente al personal laboral y administrativo. Al personal de la Unidad de Proteómica del CIB, por los conocimientos aportados y por su paciencia a la hora de analizar unas muestras tan ‘raras’. A Dr. Javier por su generosidad a la hora de proporcionarme ayuda en los ensayos de expresión de proteínas. Al Ministerio de Asuntos exteriores, especialmente a la Agencia Española de Cooperación Internacional y Desarrollo por haberme concedido una beca AECID.

Y por supuesto, millones de gracias a todos mis amigos Hatim, Hicham y Mehdi. Un gran trozo de esta Tesis también es de vosotros. Os agradezco muchísimo vuestra ayuda en los momentos difíciles y el haber compartido tantos buenos momentos a lo largo de todos estos años.

Por último, quisiera sobre todo expresar mi más profundo agradecimiento a mi familia. No hay palabras para expresar la enorme gratitud y emoción que tengo por el apoyo y amor incondicional de mis padres. Gracias por enseñarme todos los valores de la vida, qué querer es poder, por inculcarme la importancia de la honradez y el trabajo, por aceptarme y quererme tal como soy, por su comprensión, confianza, ánimos y paciencia ante los continuos aplazamientos (mamá, papá, por fin, la voy a leer. Esta vez es verdad, jeje), por el apoyo económico y múltiples sacrificios que han tomado para que intentase desarrollar una carrera científica. Muchas gracias también a mi hermano Zakaria por su especial cariño y apoyo, por ser mi mejor confidente, por tantas conversaciones transcendentales, por estar en los buenos y malos momentos. A mis hermanas Khaoula, Hanae y Rajae por su constante apoyo, por escucharme y animarme, por hacerme reír cuando estaba abatido, por ser las mejor hermanas del mundo

A mi tío Anas por su cariño, su motivación y el apoyo que me brinda constantemente, a mi abuela y a todos mis queridos sobrinos Zaid, Hind, Nada, Yasser, Hiba y Mounib

♡♡♡

El presente trabajo se ha llevado a cabo en el Centro de Investigaciones Biológicas del CSIC con ayuda de una beca de cooperación de la Agencia Española de Cooperación Internacional y Desarrollo (AECID) del Ministerio de asuntos exteriores y de cooperación.

ÍNDICE GENERAL

RESUMEN	i
SUMMARY	ii
ABREVIATURAS	iii
 CAPITULO 1 INTRODUCCIÓN GENERAL	 3
 1 LOS CARBOHIDRATOS	 4
1.1 ESTRUCTURA DE CARBOHIDRATOS	4
1.1.1. Los monosacáridos	4
1.1.2 Los oligosacáridos	6
1.1.3 Efecto anomérico y exoanomérico	7
1.1.4 Equilibrios de anomerización y de mutarrotación	9
1.2 LA IMPORTANCIA BIOLÓGICA DE CARBOHIDRATOS	9
1.2.1 Función estructural	11
1.2.2 Función de reconocimiento molecular	12
1.3 LAS INTERACCIONES ENTRE CARBOHIDRATOS Y PROTEÍNAS	12
1.3.1 Enlaces de hidrogeno	13
1.3.2 Interacciones CH- π	15
1.3.3 Las interacciones polares	16
1.3.4 El papel del agua	16
2 LAS LECTINAS	17
2.1 CLASIFICACIÓN DE LAS LECTINAS	19
2.2 LOS SITIOS DE UNIÓN DE LECTINAS	22
2.2.1 Las lectinas que reconocen galactosa	24
2.2.2 Las lectinas que reconocen N-acetilglucosamina	26
3 INTRODUCCIÓN AL ESTUDIO DE INTERACCIONES PROTEÍNA-CARBOHIDRATO MEDIANTE RMN	27
3.1 MÉTODOS DE RMN BASADOS EN LA OBSERVACIÓN DEL LIGANDO	29
3.1.1 Diferencia de Transferencia de Saturación (STD)	30
3.1.2 Experimentos de medidas de tiempo de relajación T1	32
3.2 MÉTODOS DE RMN BASADOS EN LA OBSERVACIÓN DE LA PROTEÍNA	32
3.2.1 Las perturbaciones de desplazamientos químicos del receptor	32
4 OBJETIVOS	34

5 REFERENCIAS BIBLIOGRÁFICAS	38
CAPITULO 2 MATERIALES Y MÉTODOS	44
1 MATERIALES	47
1.1 SECUENCIAS Y PROTEINAS	47
1.2 CARBOHIDRATOS	48
2 MÉTODOS	49
2.1 TÉCNICAS DE BIOLOGÍA MOLECULAR	49
2.1.1 Clonaje y transformación del DNA plasmidico	49
2.1.2 Electroforesis	49
2.1.3 Expresión y purificación del péptido recombinante	50
2.1.4 Purificación del péptido recombinante	51
2.1.4.1 Columna de afinidad de níquel	51
2.1.4.1 Cromatografía de filtración en gel	52
2.1.5 Método de Bradford	52
2.2 TÉCNICAS DE RESONANCIA MAGNÉTICA NUCLEAR	53
2.2.1 Preparación y adquisición de las muestras	53
2.2.2 Procesado de los espectros	56
2.2.3 Aplicación de RMN en la determinación de la estructura 3D de polipéptidos y proteínas	57
2.2.4 Aplicación de RMN en el reconocimiento molecular carbohidrato/proteína	64
2.2.4.1 Métodos basados en la observación del ligando	64
2.2.4.2 Métodos basados en la observación del receptor	68
2.3 TÉCNICAS DE MODELIZACION MOLECULAR	69
2.3.1 Estudio de acoplamiento o "docking" con AutoDock	69
3 REFERENCIAS BIBLIOGRÁFICAS	73

CAPITULO 3: MODELOS DE INTERACCIONES DE PROTEÍNAS CON CARBOHIDRATOS DERIVADOS DE GlcNAc	75
1 DOMINIO HEVEÍNA	79
1.1 Introducción y antecedentes	79
1.2 Resultados y Discusión	83
1.2.1 Estudio de interacción de la N,N',N''triacetilquitotriosa con la WGA	84
1.2.2 Estudio de interacción del WGA con el GMDP	89
1.3 Conclusiones	96
2 DOMINIO LYSM	97
2.1 Introducción y antecedentes	97
2.2 Resultados y Discusión	102
2.2.1 Secuencias de dominios LysM	102
2.2.2 Expresión & purificación de dominios LysM	105
2.2.2.1 Construcciones derivadas de MltD de E. coli (CAM-1-3)	105
2.2.2.2 Construcciones derivadas de la Lyk3-LysM-2 (de 1 a 7)	107
2.2.3 Caracterización estructural del dominio LysM-CAM-1	108
2.2.4 Estudios de interacciones entre LysM-CAM1 y carbohidratos	110
2.2.4.1 Interacción con quitina y derivados	110
2.2.4.2 Interacción con peptidoglicano y derivados	114
2.3 Conclusiones	120
3 DOMINIO PASTA	121
3.1 Introducción y antecedentes	121
3.2 Resultados y Discusión	126
3.2.1 Expresión y purificación del dominio	126
3.2.2 Determinación de la estructura 3D del dominio recombinante PASTA	126
3.2.3 Asignación de las señales del esqueleto	129
3.2.4 Modelización molecular de la estructura 3D del dominio PASTA-3 recombinante de PrcK de B. subtilis, mediante el programa CYANA (2.0)	134

3.2.5 Estudios de interacción entre el dominio PASTA-3 recombinante de PrkC de B. subtilis y varios carbohidratos	138
3.3 Conclusiones	140
4 CONCLUSIONES PARA EL CAPITULO 3	141
5 REFERENCIAS BIBLIOGRAFICAS	142
CAPITULO 4: CARACTERIZACIÓN DEL RECONOCIMIENTO DIFERENCIAL DE LOS ANÓMEROS DE GALACTOSA POR LECTINAS. EL CASO DE LA VISCUMINA	151
1 INTRODUCCIÓN Y ANTECEDENTES	151
1.1 La viscumina: una lectina vegetal que reconoce galactósidos	152
1.2 Hacia la diferenciación de anómeros: los equilibrios de unión frente a un receptor en el caso de sustratos mutuamente competitivos que están en equilibrio	157
2 RESULTADOS	158
2.1 Determinación de la constante de disociación de los ligandos con la VAA mediante experimentos STD	159
2.2 Experimentos de competición entre α y β metil galactósidos	177
2.3 Elaboración de mapas de epítomos por STD	182
2.4 Cálculos de "Docking" y CORCEMA	190
3 DISCUSION	198
4 CONCLUSIONES PARA EL CAPITULO 4	201
5 REFERENCIAS BIBLIOGRAFICAS	202

ABREVIATURAS

1D	Monodimensional
2D	Bidimensional
3D	Tridimensional
Å	Ángstrom
BMRB	<i>Biological Magnetic Resonance Bank</i>
CARA	<i>Computer Aided Resonance Assignment</i>
CORCEMA	<i>CORrelaxation and CONformational EXchange Matrix Analysis</i>
COSY	<i>CORrelated SpectroscopY</i>
CYANA	<i>Combined assignement and dYnamics Algorithm for NMR Applications</i>
Da	Dalton
DAP	Diaminopimelico
EC-PrkC.	<i>Extra Celular Protein Kinase C</i>
EDTA	Ácido etilendiaminotetraacético
FID	<i>Free Induction Decay</i>
Fuc	Fucosa
Gal	Galactosa
GalNAc	N-Acetilgalactosamina
Glc	Glucosa
GlcNAc	N-acetilglucosamina
GMDP	Glucosaminil-Muramil-DiPéptido
HSQC	<i>Heteronuclear Single Quantum Coherence spectroscopy</i>
Hz	Hercios
K_d	Constante de equilibrio de disociación
kDa	Kilo Dalton
k_{off}	Constante de velocidad de la reacción de disociación de un complejo
k_{on}	Constante de velocidad de la reacción de formación de un complejo
LPS	Lipopolisacáridos
Lys	Lisina

LysM	<i>Lysin Motif</i>
Man	Manosa
ML-I	<i>Mistletoe Lectin I</i>
NeuNAc	N-Acetilneuramínico
NOE	<i>Nuclear Overhauser Effect</i>
NOESY	<i>Nuclear Overhauser Effect Spectroscopy</i>
PASTA	<i>PBP And Serine/Threonine kinase Associated domains</i>
PBP	<i>Penicillling Binding Proteins</i>
PBS	Tampón fosfato salino
PDB	<i>Protein Data Bank</i>
PGN	Peptidoglicano
ppm	Partes por millón
RMN	Resonancia Magnética Nuclear
RMSD	<i>Root Mean Square Deviation</i>
SDS	Dodecilsulfato sódico
STD	<i>Saturation Transfer Difference</i>
TOCSY	<i>TOTAL Correlated Spectroscopy</i>
VAA	Viscumina
WGA	<i>Wheat Germ Agglutinin</i>
δ	Desplazamiento químico

RESUMEN

En esta tesis doctoral, se han estudiado diversas familias de compuestos proteicos (proteínas o dominios de proteínas) y de carbohidratos de distinta naturaleza química, derivados de los glicoconjugados de las paredes celulares de microorganismos mediante una combinación de métodos de espectroscopía de Resonancia magnética nuclear (RMN) y modelado molecular, para extender el conocimiento sobre el origen de la especificidad y estabilidad de complejos carbohidrato-proteína. En el tercer capítulo de este trabajo ha sido posible definir los epítomos de unión en el caso de fragmentos de peptidoglicano frente a la aglutinina de germen de trigo mediante el uso de técnicas de diferencia de transferencia de saturación RMN (desde el punto de vista del ligando). Por otro lado, en el caso del dominio LysM, se ha confirmado mediante experimentos de titulación por RMN (desde el punto de vista del ligando) la interacción del dominio LysM derivado de la transglicosilasa lítica de *Escherichia. coli* con el peptidoglicano de la propia bacteria, mientras que en el caso del dominio PASTA, ha sido posible determinar su estructura 3D por RMN sin embargo, el análisis de los experimentos de interacción (desde el punto de vista de la proteína) indicó que un único dominio PASTA aislado no interacciona con sus potenciales ligandos.

En el cuarto capítulo de esta tesis, usando como modelo la lectina de muérdago, viscumina, se ha puesto a punto una estrategia basada en RMN para estudiar de manera diferenciada cada uno de los anómeros de galactosa superando la dificultad de estudiar y analizar una mezcla de ligandos en equilibrio, indistinguibles por otras técnicas. Los resultados obtenidos en este capítulo han permitido calcular las constantes de afinidad y sacar los epítomos de unión de los ligandos en los dos sitios de unión de la viscumina a partir de las intensidades de saturación STD. Mediante la combinación de los resultados experimentales con modelado molecular y de análisis con el protocolo CORCEMA se confirma una preferencia de unión de los dos anómeros de la galactosa al sitio $\gamma 2$ (Y249) de la viscumina con una mayor selectividad hacia el anómero α . La información generada en esta tesis será de utilidad para comprender el papel que las interacciones carbohidrato-proteína juegan en algunos mecanismos utilizados por diversas células para reconocer y defenderse de patógenos y microorganismos en general.

SUMMARY

Carbohydrates, together with proteins, nucleic acids and lipids, are essential molecules of biological systems involved in structural, metabolic and cell signaling functions. Carbohydrates are composed of polyhydroxylated monosaccharide units with an aldehyde or ketone functional group. Chemically monosaccharides exist in cyclic structures closed with an acetal or hemiacetal function. Depending on the stereochemistry and regiochemistry of the acetal function, the same monosaccharide may exist in various anomeric forms. Monosaccharides generally exist in aqueous solutions in the equilibrium form between two major anomeric forms, alpha and beta. The unique reactivity of the acetal function allows monosaccharides to build oligomeric structures by forming the glycosidic bond between the anomeric position of a monosaccharide and one of the hydroxyls of the following monosaccharide resulting oligo- and polysaccharide in both lineal and branched structures of great complexity and diversity. These oligosaccharide structures can also be conjugated to other biomolecules to render glycoproteins, glycolipids, proteoglycans, etc., and in many cases are the main component of the cell outer surface (glycocalyx, polysaccharide capsules etc) or even constitute the chemical backbone of the cell walls as peptidoglycan in bacteria or chitin in fungal microorganism. Glycoconjugates are under active research area because, in addition to its structural role, have been found participating in many fundamental biological processes, fertilization, development and differentiation, immune system, recognition of pathogens, cancer metastasis and so on. Many of these phenomena involve processes of molecular recognition between carbohydrates and glycoconjugates, expressed on cell surfaces and specific carbohydrate binding proteins including lectins. Lectins are defined by their ability to bind carbohydrates but are not antibody-like proteins of immune system neither have enzymatic activity.

Carbohydrate-protein recognition processes implied transient noncovalent interactions with middle to low affinity in biochemical terms (millimolar or micro-order) but high selectivity when it is referring to distinguish different types of carbohydrates. Lectins are able to very selectively discriminate subtle differences as the stereochemistry of a single hydroxyl group that differentiates galactose from glucose. Significantly different stereochemistry of the anomeric position is what differentiates the alpha and beta anomers of monosaccharides however in this case the chemical properties of the acetal function causes the chemical equilibrium of both anomers in aqueous solution and

therefore to analyze the selectivity towards each anomer individually is not an easy task and in most cases the affinity studies are conducted on mixtures of anomers that only allow to draw average values for the equilibrium mixture but do not allow to establish preference towards anyone of the two anomers.

Methodologies based on nuclear magnetic resonance are very suitable to study carbohydrate-protein transient interactions. Both approaches are possible, observing the NMR signals of the protein spins and its perturbations by the presence of ligands or the other way around, observing the NMR signals of the ligand and its perturbations by the presence of the receptor. When applying strategies from the point of view of the protein then it is necessary to previously assign the protein's signals to the corresponding atoms of the protein. The assignment can be performed by means of homonuclear or heteronuclear correlation NMR experiments. In the case of heteronuclear experiments, mandatory for proteins over 10kD, it is necessary to express and produce, by means of molecular biology methodologies, the protein isotopically labeled with ^{13}C and/or ^{15}N to surpass the very low sensitivity of NMR. Once the chemical shift are assigned to each protein's proton then it is possible to determine the tridimensional structure of the protein recording nuclear Overhauser effects (NOE) correlations among the protons of the protein. Those NOE corresponds to dipolar interactions between hydrogen pair that are inversely proportional to the six power of interprotein distances. The interproton NOE signals can be translated into distance restrictions that become into experimental inputs for the calculations of the tridimensional structure of the by means of molecular modeling protocols. Once the structure is calculated then the perturbations of the protein NMR signals because of the interactions with the ligands can be use to estimate the affinity constants performing titrations with increasing concentration of ligand and additionally, the identification of the most perturbed signals allows to map the carbohydrate binding site in the protein. Alternatively applying NMR experiments from the point of view of the ligand make use of the difference of in the ligand's NMR parameters on passing from the small molecule regime with fast tumbling correlated with fast rotational correlation time and slow spin relaxation to the large molecule regime when the ligand is bound to the protein with slow rotational correlation time corresponding now a fast spin relaxation. Making use of these differences it is possible determine affinity constants and map the binding epitope on the ligand with NMR experiments like the saturation transfer difference (STD). STD

experiments are based on the selective transfer of spin saturation from the protein when it is selectively irradiated (without affecting directly the signals of the ligand) and this saturation is efficiently transferred only to spins that are in close contact with the protein as it happens only when the ligand is bound to the protein. relaxation. This difference between the free and the bound state amplifies the signal perturbations on the ligand spectrum and the saturation of the spins of the protein will be easily transferred to those protons of the ligand in close contact with the protein in the bound state.

In this thesis two different model systems of carbohydrate-protein interactions have been studied.

Lectins that recognize N-acetylglucosamine (GlcNAc) related oligosaccharides.

N-AcetylGlucosamine (GlcNAc) derived oligo- and polysaccharides participate in many biologically important structures. Chitin (GlcNAc polysaccharide) is the main structural component of insect and crustaceans exoskeleton and is present in the cell wall of fungus and other microorganism, as well as bacterial cell walls are made of peptidoglycan that has its glycan core chain built with disaccharide repeating units of GlcNAc and related N-AcetylMuramic acid (GlcNAc-N-Acetylmuramyl disaccharides) crosslinked with short polypeptide fragments. Quite diverse proteins recognize those chito-oligosaccharide related structures accomplishing functions as, for example, defense proteins (hevein domains), symbiosis signal recognition (Lys-M domains), carbohydrate processing (chitinases, glucosaminidases) etc.

In this thesis NMR methodologies have been applied for studying those interactions in model systems based on hevein, Lys-M and PASTA protein domains.

-Hevein Domains:

Many plants express chitin binding proteins that have been identified as plant defense proteins. A family of these plant defense proteins contains one or several chitin binding sites localized on the so called "hevein domains". Those domains have been named after hevein, a small 43 residues protein with 4 disulfide bonds that binds chitin and chito-oligosaccharides.

Wheat Germ agglutinin (WGA) is a lectin composed of four homologous hevein domains. It has been shown that WGA binds N-AcetylGlucosamine (GlcNAc) related oligosaccharides and in fact it is used as commercial reagent to detect glycans containing

GlcNAc residues. It has been shown that WGA is able to bind bacterial cells, data that could correlate with its plant-defense capacities, but there is no information at molecular level how WGA binds to peptidoglycan. In this thesis the binding of short peptidoglycan fragments to WGA have been characterized by means of STD-NMR studies. The results obtained demonstrate that disaccharide dipeptide peptidoglycan fragment (glucosaminyl-muramyl-dipeptide, GMDP) is recognized by WGA and that the GlcNAc residue establishes the mayor contacts with WGA followed by the N-Acetylmuramic acid residue while the peptidic moiety is pointing outside of the binding site. Additionally, as the GMDP is present as the mixture of anomers of its reducing end (muramic residue), the STD experiments have shown that WGA preferentially binds the the beta anomer.

-LysM-Domains

The LysM (Lysin motif) named after *Escherichia c.* lytic transglycosidase is a protein domain of about 50 aminoacid residues being part of larger multidomain proteins extensively distributed in virus, bacteria, plants, fungi and animals. In bacteria and plants some of the proteins containing LysM domains are associated with recognition of peptidoglycan and quitoologosaccharides acting as sensor proteins of pathogens (recognition of patogen by binding fragmentes of the cell walls of the patoghens) or simbionts (for example the receptors of bacterial nodulation factors (chitolipooligosaccharides) contains three LysM domain sequentially organized). However there is not yet a generalized description of the binding mode between oligosaccharides and LyM domain. In this thesis a mutant variant of LysM domain from *Escherichia c.* lytic transglycosidase has been used as model system to study LysM-oligosaccharide interactions. First the domain was successfully expressed and produced from bacteria. The attempts to similarly express some LysM domains from plant proteins involved in detecting nodulation factors were unsuccessful. The correct folding of the model LysM domain was confirmed by assignment of the homonuclear NMR spectra of the domain. After assignment and structural calculations, binding studies were performed following possible perturbation of signals from the protein. Only with polymerized peptidoglycan was possible to shown binding of LysM-domain to peptidoglycan following NMR signal intensity decrease after several additions of peptidoglycan to the sample however no interaction was observed neither with GMDP nor with quitoologosaccharides.

-PASTA Domain:

Reactivation from dormancy, growth and division of bacteria require cleavage of the cell-wall peptidoglycan. Bacteria turn over their cell-wall material due to the actions of peptidoglycan hydrolases and amidases

Muropeptide-driven exit from dormancy requires a member of the serine/threonine kinase (STPK) family. Proteins of this family are expressed in many prokaryotes, including a broad range of pathogens, and modulate a wide number of cellular processes, such as biofilm formation, cell wall biosynthesis and cell division, sporulation, stress response.

This proteins presents in their domain organizations, an intracellular serine/threonine kinase domain, a transmembrane region and an extra-cellular portion that contains domains denoted as PASTA (*Penicillin binding Associated and Serine/Threonine kinase Associated domains*) It has been described that the PASTA domain is responsible of the binding to muropeptides. In this thesis it has been determined by NMR, applying heteronuclear NMR correlations experiments, the tridimensional structure of the second PASTA domain of Prkc protein from *Bacillus subtilis* expressed heterogously with isotopic enrichment in ^{13}C and ^{15}N . Interestingly, when the isolated domain was studied in the presence of muropeptides isolated from *Bacillus s.* cells walls, no interaction was observed. Additionally no binding was observed with betalactams neither with GMDP muropeptide. The results obtained rise the hypothesis that more than one PASTA domain should be necessary to express the full functionality of this protein domains in order to recognize their natural ligands.

Lectins that binds Galactose. Anomers selection by viscumin

Viscumin is plant toxin-type lectin extracted from latex of mistletoe and has been proposed to be useful therapeutic drug for treatment some types of cancer. It is a heterodimer with a toxin subunit A and a lectin subunit B. The lectin subunit has two carbohydrate binding sites, one located around tryptophan 38 y other around tyrosine 249. The tryptophan site has been characterized as the one with high affinity but its functionality is very much dependent of the quaternary structure of the lectin. Previous work have shown some preference for alpha galactosides but there is no data if the viscumin discriminate the anomers of plain reducing galactose monosaccharide.

Reducing carbohydrates can be characterized in two different structures depending of the configuration at the anomeric position of the reducing end. Anomeric selectivity has been observed and studied in enzymes involved in carbohydrate transformations by diverse kinetic approaches since long time ago, however those approaches are not possible in the case of just binding studies of carbohydrate in solution. Lectins and other carbohydrates receptors do not modified their ligands, thus do not generate time dependent responses feasible of monitoring. Taking into account that free anomers equilibrate in no more than a few hours, many of the available biophysical techniques in solution only provide a mean macroscopic view of the recognition event involving the mixture of anomers in equilibrium. The direct discrimination of the recognition of each anomer is not an easy task, and NMR can be technique of choice because it allows observing independent signals for each anomer in spite of they are in chemical equilibrium exchange. In fact, ^{13}C NMR has been used to show the α -anomer preference of bacterial chemotaxis sugar binding proteins and more exotic tritium ^3H -NMR has been used to study anomeric preference in maltose binding protein. In this thesis, by means of STD-NMR strategies combined with molecular docking, it has been possible to obtain quantitative information on the individual interaction of each anomer, in order to characterize the differential monosaccharide anomer affinity by viscumin lectin. the obtained data shows an slight preference for the α anomer over the β anomer in both cases, the free reducing galactose as mixture of anomers and the α - and β - methyl-galactosides with fix configuration at the anomeric carbon. On the other side the comparison of the experimental STD data with theoretical predictions of saturation for each proton of the ligand applying the CORCEMA protocol allows to define the binding epitope of each ligand and identify the binding site on the protein. Interestingly, under the conditions of viscumin concentration used in the NMR experiments that favors the dimer of heterodimers ABAB quaternary structure, the fitting of the experimental with the calculated STD data shows that the tryptophan site of viscumin is not full accessible to the ligand while the experimental data can be explained assuming that the ligand binds mostly to the tyrosine site of viscumin. This result is in accordance with previous reports that shown the dependence of the functionality of the tryptophan site on the viscumin concentration.

CAPÍTULO 1

INTRODUCCIÓN GENERAL

Capítulo 1: Introducción General

En la presente tesis se presentan estudios, mediante técnicas de espectroscopia de Resonancia Magnética Nuclear (RMN) en disolución, de procesos de reconocimiento molecular de carbohidratos relacionados con polisacáridos de paredes celulares, tanto por la parte de lectinas que reconocen N-acetilglucosamina y galactosa, específicamente la WGA y la Viscumina como por otros dominios proteicos llamados LysM y PASTA, con el objetivo de la elucidación de las bases moleculares de estos procesos y el origen de su selectividad.

Previamente a la presentación de los resultados obtenidos, en primer lugar, se introducirán las bases moleculares de las interacciones carbohidrato/proteína en las que se centra esta Tesis, y, seguidamente se describirán las técnicas de RMN que se han usado a lo largo de este trabajo de investigación.

1 LOS CARBOHIDRATOS

La glicociencia estudia la estructura, propiedades y funciones de los carbohidratos en la materia viva [1]. Estos estudios no se refieren sólo a los carbohidratos, sino también a todas otras clases de moléculas biológicas (proteínas, lípidos y ácidos nucleicos) que interactúan con ellos. Esta primera parte está dedicada al estudio estructural de los carbohidratos y de las interacciones que pueden establecer.

1.1 ESTRUCTURA DE CARBOHIDRATOS

1.1.1 Los monosacáridos:

Los hidratos de carbono (carbohidratos o azúcares) están constituidos formalmente, como su nombre indica, a partir de átomos de carbono y moléculas de agua- Se representan por la fórmula empírica $C_n(H_2O)_n$ y se producen en la naturaleza como tal, o también en forma conjugada con las proteínas y lípidos (glicoconjugados). En estos glicoconjugados, la parte del carbohidrato puede ser responsable de su solubilidad, estabilidad y sus interacciones.

La D-glucosa ($C_6H_{12}O_6$) es el monosacárido más común. Es la principal fuente de energía para la mayoría de las células de organismos superiores y se puede describir químicamente en tres formas diferentes: una estructura lineal y dos estructuras cerradas por una función hemiacetal, que forman anillos de 6 ó de 5 átomos (piranosa y furanosa respectivamente) (Figura 1.1). La forma predominante en disolución acuosa de la glucosa es la forma piranosa. Cuando se cicla, se crea un nuevo centro de asimetría respecto a la estructura lineal. De esta manera, se pueden formar dos estructuras: la α -D-glucopiranosa y la β -D-glucopiranosa. Al C-1 se le denomina carbono anomérico y las formas α y β son anómeros. Hay que resaltar que todas estas formas, que responden a la misma fórmula molecular, son estructuras químicas distintas con estabilidades relativas diferentes. Dependiendo del medio en que se encuentren, se favorecen unas formas sobre otras. Por ejemplo, la glucosa cristaliza preferentemente en la forma α -D-glucopiranosa. Igualmente, esta forma es la más estable en estado gaseoso mientras que en disolución acuosa, la forma β -D-glucopiranosa está más favorecida. El anillo piranosa de seis eslabones no puede ser plano, debido a la geometría tetraédrica de los átomos de carbono saturados. De hecho, los anillos de piranosa pueden adoptar diversas

conformaciones (silla, *C chair*; bote, *B boat*; bote torcido, *S skew boat*; y media silla, *H half chair*), siendo las más estables las denominadas silla. Los anillos de furanosa tampoco son planos. En general, se doblan de modo que tres o cuatro de sus átomos son coplanares [2].

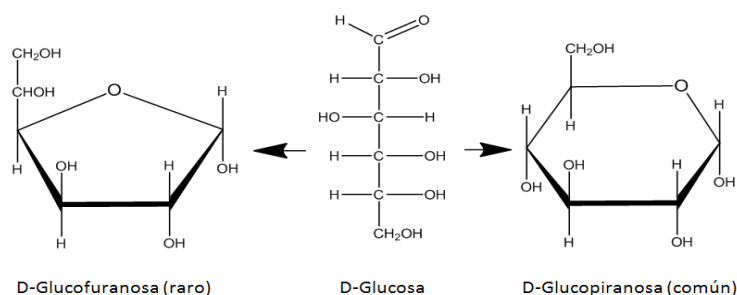


Figura 1.1 Las tres formas del D-glucosa, se generan a partir de la molécula lineal por reacción del aldehído en el carbono 1 con el hidroxilo en el carbono 5 o carbono 4, la forma de piranososa (derecha) predominante en los sistemas biológicos.

Como regla general, los sustituyentes voluminosos suelen adoptar una disposición pseudo-ecuatorial para evitar interacciones 1,3-diaxiales. Así, en el caso de la D-glucosa, la conformación 4C_1 , con todos sus sustituyentes ecuatoriales, es mucho más estable que la conformación 1C_4 , en la que todos son axiales [3] (Figura 1.2).

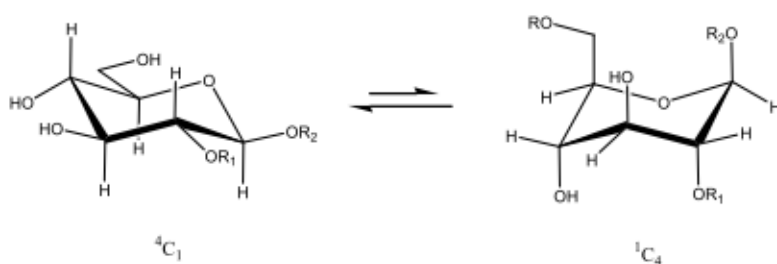


Figura 1.2 Representación de las conformaciones 4C_1 y 1C_4 de la D-glucosa.

Hasta ahora hemos descrito a los carbohidratos con la fórmula molecular $C_n(H_2O)_n$, pero dentro de esta familia de compuestos se incluyen también todos aquellos derivados de los azúcares que muestran modificación o sustitución con otros grupos funcionales (Figura 1.3). Así, se pueden mencionar los aminoazúcares en los cuales se ha sustituido algún grupo hidroxilo por un grupo amino o acetamido. Por ejemplo, glucosamina,

galactosamina. y sus variantes acetiladas N-acetilglucosamina, N-acetilgalactosamina, etc. También existen azúcares deshidroxilados, como la fucosa (Fuc) (o 6-desoxigalactosa) y la desoxiribosa, y otros que presentan algún grupo carboxilo, como los ácidos urónicos y los ácidos siálicos. Finalmente, pueden mencionarse aquellos que sufren otras modificaciones, como sulfatación o metilación de los grupos hidroxilo. Todas estas modificaciones hacen que los carbohidratos, incluso al nivel de sus unidades monosacáridicas más sencillas, presenten una diversidad química muy extensa.

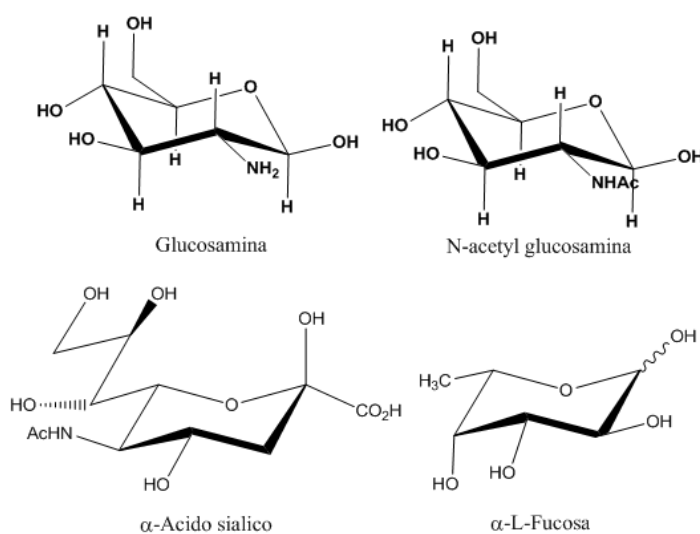
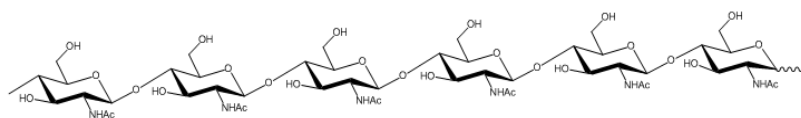


Figura 1.3 Ejemplos de carbohidratos que sufren modificación o sustitución de grupos funcionales.

1.1.2 Los oligosacáridos

Los monosacáridos se pueden unir entre sí formando disacáridos, mediante un enlace de tipo acetálico denominado enlace glicosídico. Aquí, el hidroxilo libre de la función hemiacetálica del primer monosacárido, que se denomina resto glicosilo o extremo no reductor, se sustituye por un hidroxilo del otro monosacárido. Este se denomina extremo reductor. Este tipo de uniones glicosídicas permite formar desde los más simples disacáridos hasta oligosacáridos, tanto en estructuras secuenciales como ramificadas (Figura 1.4). Las estructuras así formadas pueden ser de gran complejidad, clasificándose como polisacáridos cuando contienen diez o más unidades monosacáridicas.

A)



B)

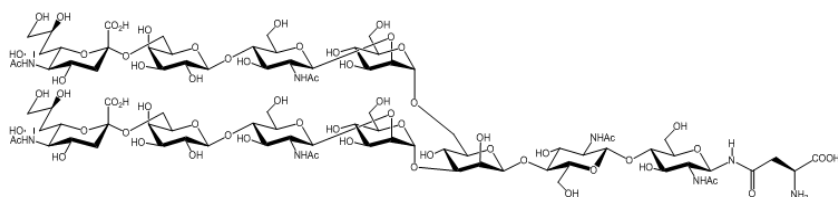


Figura 1.4 Ejemplo de una estructura lineal (A) y ramificada (B) de oligosacáridos.

Para la caracterización estructural de los oligosacáridos, se requiere definir al menos dos ángulos de torsión ϕ (Φ) y ψ (Ψ). Así, se puede detallar la orientación relativa de las unidades de monosacárido (Figura 1.5)

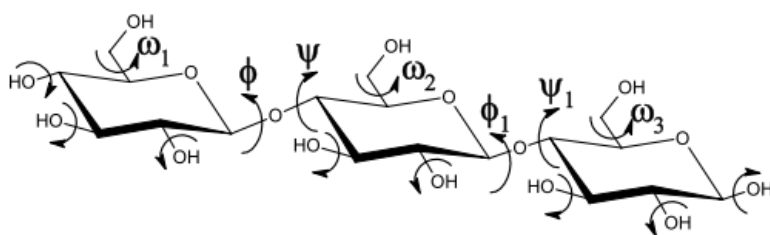


Figura 1.5 Los ángulos de torsión glicosídicos Φ y Ψ para el trisacárido $\text{Glc}\beta 1 \rightarrow 4\text{Glc}\beta 1 \rightarrow 4\text{Glc}$.

Los ángulos de torsión (Φ , Ψ) se definen como $\Phi = \text{H1-C1-O1-C}_x$ y $\Psi = \text{C1-O1-C}_x\text{-H}_x$, según nomenclatura de resonancia magnética nuclear (RMN) ($\Phi = \text{O5-C1-O1-C}_x$ y $\Psi = \text{C1-O1-C}_x\text{-C}_{(x+1)}$, según la nomenclatura cristalográfica). En concreto, C_x se refiere a la posición del residuo más próximo al extremo reductor. Cuando la unión glicosídica implica el grupo hidroximetilo (posición 6 en hexopiranosas), se requiere definir otro

ángulo diedro llamado omega (ω) para definir la orientación espacial relativa de los residuos constituyentes del oligosacárido ($\omega = \text{O}_{5x}-\text{C}_{5x}-\text{C}_{6x}-\text{O}_{6x-1}$).

1.1.3 Efectos estereoelectrónicos: efecto anomérico y exoanomérico

Las propiedades estereoelectrónicas de la función hemiacetálica/hemicatálita en monosacáridos o acetálica/cetálica en oligosacáridos son un factor determinante de la estructura tridimensional de los carbohidratos y son el origen de los denominados efectos anomérico y exoanomérico. [3-6].

El efecto anomérico determina que, para los sustituyentes electronegativos en posición 1 del anillo de piranosa, la orientación axial (configuración α en la serie D es más estable que la ecuatorial (configuración β en la serie D) [7]. En contra de lo previsible, esta preferencia se produce debido a la deslocalización hiperconjugativa de un par de electrones libres del oxígeno del anillo de piranosa con el orbital σ^* del enlace contiguo C1-O1. Esta deslocalización únicamente es posible en la orientación axial (Figura 1.6). Este efecto tiene una gran importancia en las moléculas que tienen dos heteroátomos unidos a un centro tetraédrico y tiene también implicaciones geométricas. Por ejemplo, en la configuración axial (anómero α) se observa un acortamiento del enlace O5-C1, un alargamiento del enlace C1-O1 y un aumento del ángulo O5-C1-O1, con respecto al análogo ecuatorial.

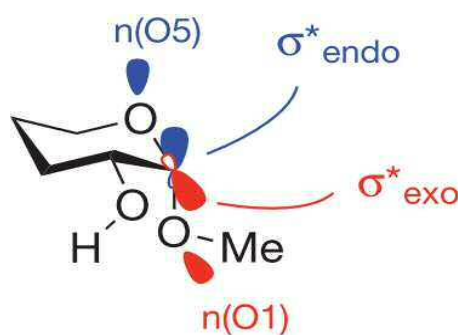


Figura 1.6 Representación esquemática de las de las interacciones responsables del efecto anomérico $n(\text{O}_5) \text{Rs}^*(\text{C1}-\text{O1})$ (azul) y exoanomérico $n(\text{O1}) \text{Rs}^*(\text{C1}-\text{O5})$ (rojo).

Por otro lado, el efecto exanómico se refiere al predominio de una orientación de tipo *gauche* para la torsión exocíclica. Por tanto, es un parámetro de gran importancia para definir la conformación de oligosacáridos. Se explica, igualmente que el efecto anómico, mediante la teoría de orbitales moleculares.

El efecto exanómico afecta al ángulo de torsión definido por el enlace glicosídico. De hecho, determina que se encuentren favorecidas aquellas conformaciones alrededor del enlace inter-glicosídico que permiten la interacción entre el par electrónico del átomo de oxígeno unido a la posición 1 del anillo de piranosa con el orbital σ^* del enlace endocíclico C1-O5 (Figura 1.6).

1.1.4 Equilibrios de anomerización y de mutarrotación

Como se ha mencionado, la preferencia entre las distintas formas anómicas en monosacáridos depende del efecto anómico y del medio. En el contexto de los sistemas biológicos es muy importante resaltar que los azúcares disueltos en medios acuosos sufren una interconversión entre las distintas formas anómicas en una escala de tiempo de minutos-horas (dependiendo de las condiciones de pH, temperatura, y presencia de catalizadores) mediante el proceso de mutarrotación. El estado de equilibrio correspondiente entre las distintas formas anómicas se alcanza finalmente. Por ejemplo, para la *D*-glucosa en disolución acuosa, el equilibrio contiene aproximadamente un 62% β y 37,6 % α , junto a proporciones muy minoritarias de las formas furanosas y abierta [8] (Figura 1.7).

La importancia de la presencia de ambas formas anómicas en medios biológicos se pone de manifiesto por la existencia de enzimas mutarrotasas [9] y por la distinta selectividad observada hacia los distintos anómeros en diversos procesos enzimáticos implicados en el metabolismo de carbohidratos, estudiados mediante medidas cinéticas de polarografía o RMN [10-11]. Por el contrario, la caracterización del reconocimiento diferencial de anómeros por lectinas (receptores de carbohidratos) en solución ha sido poco abordada, dado que las lectinas unen los carbohidratos sin modificarlos y, por tanto, no generan respuestas dependientes del tiempo que puedan ser monitorizadas.

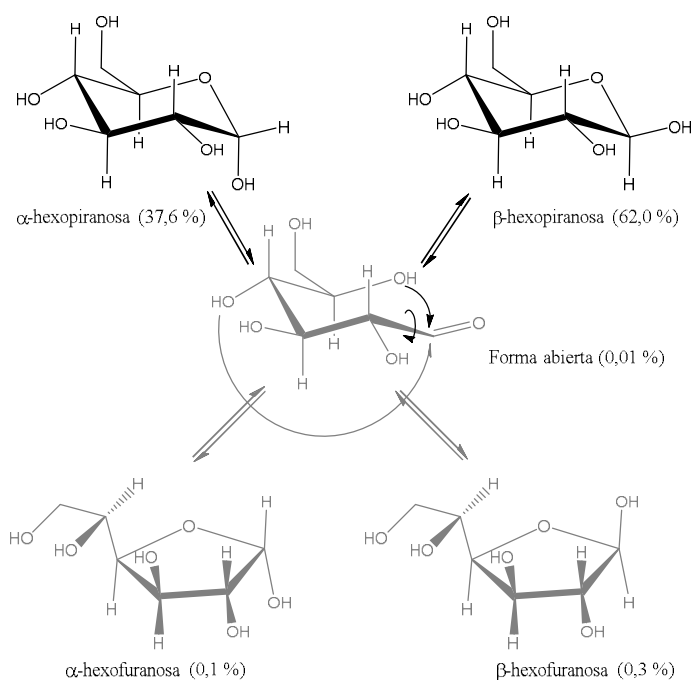


Figura 1.7 Equilibrio de mutarrotación para la D-glucosa (porcentajes en equilibrio en agua a 30°C).

Teniendo en cuenta que los anómeros en disolución se equilibran en unas pocas horas, la mayoría de las técnicas biofísicas disponibles únicamente permiten obtener valores promedio (macroscópicos) en los procesos de reconocimiento que implican mezclas de anómeros. Cuando se trata de estudiar independientemente el comportamiento de un anómero específico, se recurre habitualmente a una estrategia química, modificando la posición anomérica (por ejemplo mediante metilación con objeto de producir la sustitución más simple posible). Así, se rompe el intercambio anomérico y se generan dos compuestos con su configuración anomérica, α o β , prefijada. Afortunadamente, la RMN permite la discriminación directa del proceso de reconocimiento de cada anómero en la mezcla. Así, por ejemplo, se ha aplicado la resonancia de ^{13}C para mostrar la preferencia por el anómero β de las proteínas de unión a azúcares implicadas en quimiotaxis bacteriana [12]. Además, estudios “más exóticos” de resonancia de tritio (^3H -NMR) se han usado para determinar la preferencia anomérica de la proteína de unión a maltosa [13]. Recientemente, en nuestro grupo de trabajo, se ha mostrado que también la resonancia de ^{19}F se puede aplicar para determinar la selectividad anomérica en un proceso de reconocimiento molecular [14].

1.2 LA IMPORTANCIA BIOLÓGICA DE CARBOHIDRATOS

Los carbohidratos son componentes esenciales de todos los organismos vivos, desde los virus y microorganismos unicelulares hasta los organismos superiores pluricelulares. Desde el punto de vista de su estructura química, muestran una gran variabilidad de composición y secuencia, y una alta diversidad geométrica acompañada de una gran flexibilidad estructural. Así, un mismo glicoconjugado puede poblar varios estados conformacionales que coexisten en equilibrio bajo ciertas condiciones fisiológicas. Estas propiedades estructurales confieren a los carbohidratos una extraordinaria capacidad de contener y transmitir información biológica específica [15]. Los carbohidratos participan, y en muchos casos son protagonistas principales, en una gran variedad de funciones biológicas. Por ejemplo, son la principal fuente de energía metabólica y el componente estructural esencial de las paredes celulares (quitina, celulosa peptidoglicano) y la matriz extracelular (proteoglicanos y glicosaminoglicanos, exopolisacáridos fúngicos y bacterianos). Conforman sitios de interacción en la superficie celular esenciales en procesos de reconocimiento intercelular, como los procesos infecciosos o de simbiosis, la metástasis, o reconocimiento inmunológico (grupos sanguíneos), de adhesión, de señalización y otros[15-16].

En todos estos procesos participan una gran variedad de glicoconjugados y sus fragmentos metabólicos (glicolípidos, glicoproteínas, proteoglicanos, lipopolisacáridos bacterianos, lipoquitoligosacáridos, fragmentos de peptidoglicano etc.). Las funciones biológicas asociadas con los carbohidratos son muy diferentes, igual que su diversidad estructural. Sin embargo, dos grandes tipos de funciones se pueden destacar: la función estructural y la función de reconocimiento molecular

1.2.1 Función estructural

La mayoría de los carbohidratos se encuentran en la naturaleza como parte de los polisacáridos. Los homopolisacáridos contienen un único tipo de monosacáridos. Son moléculas muy importantes para el almacenamiento de energía química; el almidón (para las plantas) y el glucógeno (para los animales) se componen solamente de unidades de glucosa. Desde el punto de vista estructural, la quitina forma parte de las paredes celulares de los hongos y del resistente exoesqueleto de los artrópodos. Por otra parte, los heteropolisacáridos contienen diferentes tipos de unidades monosacáridicas.

Por ejemplo, mientras que las paredes de las células de plantas se constituyen a menudo a partir de homopolisacáridos como la celulosa, las de las células bacterianas se componen de unidades heterogéneas de peptidoglicano. Estas moléculas a menudo tienen una carga electrostática alta, una gran área superficial y una alta hidratación. Estas propiedades físicas juegan un papel fundamental en la estabilización estructural, la solubilización y la protección contra la proteólisis de glicoproteínas. También forman parte de la capa de alta densidad molecular, el “glicocalix”, que cubre la superficie de las células epiteliales en eucariotas. Esta región está involucrada en la cohesión celular y en la protección de las células contra los choques físicos y el ataque de microorganismos.

1.2.2 Función de reconocimiento molecular

Los carbohidratos están involucrados en muchos de los fenómenos de reconocimiento molecular receptor-ligando que implican las lectinas y muchos estudios confirman su participación en la vida social de las células [17]. Hoy en día está claro que el principal papel biológico de los carbohidratos es el de ejercer de puntos específicos de reconocimiento molecular para lectinas, enzimas y anticuerpos. [18] (Figura 1.8)

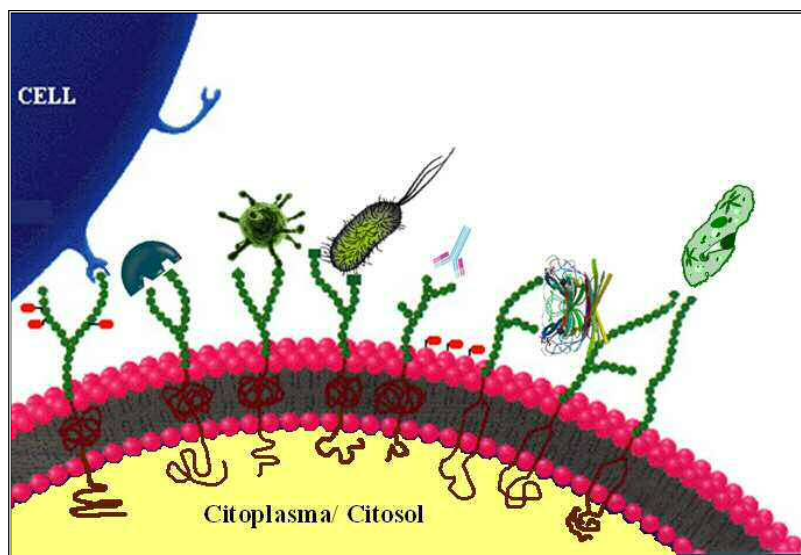


Figura 1.8 Esquema idealizado que muestra el papel de los glicoconjugados situados en la superficie celular en las interacciones en diverso grado de complejidad desde biomoléculas (lectinas, anticuerpos toxinas...) hasta células y microorganismos completos (virus, bacterias, parásitos etc.).

Las interacciones entre carbohidratos y proteínas juegan un papel esencial, mediando muchos fenómenos fundamentales y complejos. Pueden mencionarse reacciones alérgicas, la embriogénesis, el desarrollo de los tejidos, la fertilización, la metástasis [19-20], así como el reconocimiento de los epítomos de la pared extracelular y bacteriana. La mayoría de los microorganismos, patógenos o simbios, incluyendo microbios, virus, bacterias y hongos, han desarrollado estrategias de interacción mediante lectinas y toxinas para unirse a los carbohidratos de la superficie de las células diana con el fin de mantener interacciones imprescindibles para una futura infección y colonización. Además, desde el punto de vista del carbohidrato, sintetizan glicoconjugados (lipoquitooligosacáridos) como señales elicitoras en procesos de simbiosis [21] o, en sentido opuesto, glicoconjugados de sus paredes celulares (fragmentos de quitina o peptidoglicano, lipopolisacáridos...) disparan señales de defensa y activan las respuestas inmunológicas en las células diana [22-25].

1.3 LAS INTERACCIONES ENTRE CARBOHIDRATOS Y PROTEÍNAS

1.3.1 Enlaces de hidrogeno

Los carbohidratos son moléculas polihidroxiladas y, por tanto, tienen una gran potencialidad para establecer enlaces de hidrógeno tanto como donadores o aceptores. De hecho, los enlaces de hidrogeno son elementos esenciales en la interacción entre proteínas y carbohidratos [26-27]. El oxígeno de los grupos hidroxilo (-OH), con hibridación sp^3 , tiene una disposición aproximadamente tetraédrica de dos pares de electrones y un protón. Los OHs, por lo tanto, pueden actuar como un receptor de dos enlaces de hidrógeno, y como donante de uno solo [28-29]. Adicionalmente, hay que mencionar que el enlace de hidrógeno cooperativo, donde el grupo hidroxilo (OH) actúa simultáneamente como donador y aceptor de enlace de hidrógeno, es característico de la interacción de las lectinas y otras proteínas que se unen a los hidroxilos de azúcares [26]. En general, las cadenas laterales ácidas en proteínas pueden actuar como aceptores de enlaces de hidrógeno, a partir de uno o dos OHs del azúcar. Los donantes de enlaces de hidrógeno provienen principalmente de los grupos amida de la cadena principal o del grupo amida de cadenas laterales de asparagina y, menos frecuentemente, de glutamina. También, aunque es menos común, los OHs de las cadenas laterales de tirosina, serina y treonina participan como donantes o aceptores de enlaces de hidrógeno con los OHs del

azúcar [30]. De este modo, el esquema de unión del hidrógeno más común relacionado con los OHs del azúcar es de tipo cooperativo, donde el grupo hidroxilo (OH) actúa simultáneamente como donador y aceptor de enlace de hidrógeno: $(\text{NH})_n \rightarrow \text{OH} \rightarrow \text{O} = \text{C}$; donde el NH es el grupo donador del enlace de hidrógeno, $\text{O} = \text{C}$ es el carbonilo aceptor de la cadena principal de grupos carboxilato o carboxiamida, OH es un hidroxilo del azúcar, y n puede valer 1 o 2. La geometría del enlace de hidrógeno entre aceptores y donadores es fija, mientras que un OH tiene la libertad de torsión en la ubicación del protón y los pares no enlazantes. Presumiblemente, la libertad de rotación permite la optimización de los enlaces de hidrógeno entre el grupo OH y sus grupos vecinos, aunque con cierto coste entrópico, debido a los rotámeros preferentes de los grupos OH.

El átomo de oxígeno del anillo de azúcar también tiene hibridación sp^3 y tiene dos pares de electrones en la geometría tetraédrica que pueden actuar como aceptores de enlaces de hidrógeno. Sin embargo, no puede participar en los enlaces cooperativos de hidrógeno. En general, este oxígeno es común en todos los azúcares y, por tanto, no puede utilizarse para distinguir entre ellos. La presencia de otros grupos funcionales en carbohidratos como grupos carboxilo, acetamido y sulfato, amplía la diversidad de tipos de puentes de hidrógeno que pueden formar los carbohidratos en sus interacciones con proteínas. Los grupos acetamido de N-acetilglucosamina (GlcNAc), N-acetilgalactosamina (GalNAc), y N-Acetilneuramínico (NeuNAc) son un factor determinante para la selectividad en el reconocimiento de estos carbohidratos. Al contrario del grupo hidroxilo, el grupo amida y el oxígeno carbonílico del sustituyente acetamido tienen una geometría fija y plana. El grupo amida actúa como donador de enlaces de hidrógeno al carbonilo planar o a los oxígenos del carboxilato en todos los casos observados hasta la fecha, mientras que el oxígeno acetamido acepta con frecuencia los enlaces de hidrógeno por parte de la serina.

Desde el punto de vista termodinámico, hay que tener en cuenta que los enlaces de hidrógeno que se forman en la interacción carbohidrato-receptor deben, por un lado, compensar energéticamente la ruptura de los correspondientes enlaces de hidrógenos del carbohidrato con el agua del medio y, por otro, introducir los elementos de selectividad en el reconocimiento de unos carbohidratos y no de otros. En general, las funcionalidades del azúcar que forman enlaces de hidrógeno con las lectinas son las necesarias para el reconocimiento específico y la discriminación, mientras que aquellas

posiciones que no se utilizan como elementos de reconocimiento tienden a estar más expuestas al disolvente, aunque pueden formar contactos indirectos con la proteína, por ejemplo, a través de moléculas de agua (Figura 1.9).

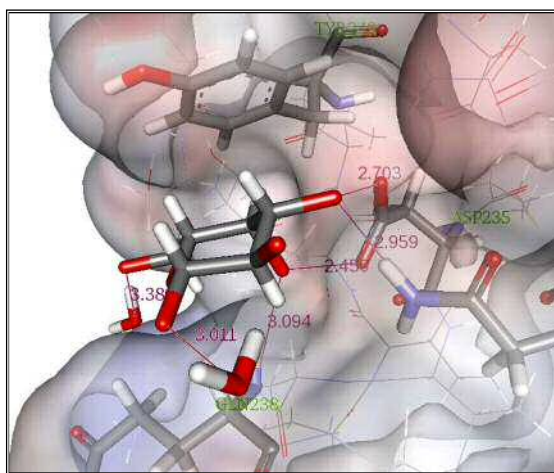


Figura 1.9 Representación de los diferentes enlaces de hidrógeno con sus correspondientes distancias en Å de la β-galactosa en el sitio de unión γ 2 de la cadena B de la viscumina (código PDB: 1PUM).

Los enlaces de hidrógeno entre las lectinas y los determinantes esenciales del reconocimiento del azúcar están habitualmente más protegidos del disolvente, lo que significa que están en un entorno dieléctrico bajo por lo que sus entalpías de unión son previsiblemente más fuertes que las formados con el agua [31-32]. Además, el uso de donadores y aceptores de enlaces de hidrógeno con geometría fija puede ser importante para la especificidad y para minimizar costes entrópicos. Por otro lado, un grupo con gran libertad de giro, como el OH de la serina, tiene cierta plasticidad/adaptabilidad para formar enlaces de hidrógeno con el azúcar [30, 33]. En este sentido, puede que no sea capaz de discriminar entre grupos hidroxilo epímeros de dos azúcares distintos.

1.3.2 Interacciones CH- π

Ya se ha mencionado que los carbohidratos son moléculas polihidroxiladas. No obstante, además hay que decir que presentan superficies no polares, que contienen enlaces C-H. Por ello, tienen un carácter anfífilico y, en función de su naturaleza química, pueden mostrar superficies no polares capaces de interaccionar con residuos hidrófobos complementarios de Tyr, Phe y Trp. De este modo, los enlaces C-H

localizados en la misma cara del carbohidrato, próximos en el espacio, pueden participar en interacciones CH- π con las nubes π de las cadenas laterales de los aminoácidos aromáticos [34-35]. Las interacciones CH- π son muy comunes en los sitios de reconocimiento de las proteínas que reconocen los carbohidratos tales como las galectinas [36] o los dominios de heveína. Este tipo de interacciones con residuos aromáticos se ha investigado ampliamente en nuestro laboratorio [37-41]. La fuerza de esta interacción es relativamente débil, ya que una única interacción CH- π es del orden de 1 kcal/mol [38, 42-43]. Además de la cara más apolar del anillo de azúcar, el grupo metilo de acetamidas en GlcNAc y NeuNAc interacciona a menudo con el anillo aromático en lectinas, como la WGA [30, 33, 44]. El uso frecuente de las cadenas laterales aromáticas en las interacciones con azúcares [45] indica que la interacción con la nube de electrones deslocalizados del anillo aromático es energéticamente significativa, más allá de proporcionar simplemente una superficie apolar geoméricamente complementaria. En el caso de la galactosa, siempre se encuentran cadenas laterales aromáticas interaccionando. Sin embargo, esta situación apenas se da con manosa. Estas observaciones pueden reflejar el hecho de que la disposición del grupo 4-OH axial en Gal crea una superficie apolar continua y muy amplia.

1.3.3 Las interacciones polares

Aparte de las interacciones por puente de hidrogeno ya mencionadas, también se encuentran casos de interacciones polares como la coordinación con iones metálicos (por ejemplo en el caso de las C-lectinas, ver más adelante figura 1.14) o interacciones electrostáticas entre residuos cargados de la proteína y oligosacáridos modificados con grupos ionizables, como por ejemplo en el caso del reconocimiento de glicosaminoglicanos donde residuos de arginina y lisina establecen interacciones electrostáticas con los grupos carboxilo y sulfato de los oligosacáridos[46-48].

1.3.4 El papel del agua

Teniendo en cuenta el carácter polihidroxilado de los carbohidratos, es evidente que sus interacciones con el medio acuoso fisiológico debe jugar un papel esencial en el balance global energético de las interacciones con sus receptores. Es necesario tener en cuenta todos los procesos de solvatación-desolvatación que tienen lugar en la formación de los

complejos carbohidrato-receptor [49]. Adicionalmente, también se ha observado que las moléculas de agua pueden jugar un papel específico al mediar en enlaces de hidrógeno entre el azúcar y la proteína, según se ha determinado en diversas estructuras tridimensionales de complejos carbohidrato-lectina [32]. Estudios de los inhibidores de glucógeno fosforilasa han demostrado que los enlaces de hidrogeno mediados por agua entre el azúcar y la proteína pueden ser tan fuertes como los enlaces de hidrógeno directos entre la proteína y el carbohidrato [50]. La comparación de una serie de azúcares unidos a una lectina dada, o una serie de lectinas unidas a un azúcar dado, revela, a veces, moléculas comunes de agua que se pueden presumir que sean elementos importantes en el reconocimiento. Por ejemplo, se ha observado que tres moléculas de agua median una misma interacción entre galactosa (Gal) con una enterotoxina (“heat-labile”) y con la toxina del cólera [51-52], así como entre la lactosa y otros receptores LT [52-53]. Además, la comparación del sitio de unión entre los estados libre y asociado muestra que algunas moléculas de agua unidas a la proteína en el estado libre de azúcar permanecen en presencia del azúcar y forman enlaces de hidrógeno con él. En este caso, las moléculas de agua actúan como elementos estructurales fijos y, por lo tanto, se pueden considerar como parte de la arquitectura del sitio de unión [54].

2 LAS LECTINAS

La palabra lectina viene del verbo latino *LEGERE* que significa “seleccionar” o “elegir”, y fue propuesta por W. C. Boyd en 1954 [55-56] para designar a las aglutininas específicas de los grupos sanguíneos extraídas de ciertas semillas vegetales. El tiempo ha venido a confirmar lo apropiado de esta denominación para esta clase importante de proteínas que “seleccionan” carbohidratos. Actualmente, las lectinas engloban a aquellas proteínas de origen no inmune que se unen específicamente y de forma reversible a los azúcares, sin mostrar actividad enzimática sobre estos [57-59].

Según describe H.-J. Gabius en su artículo “History of Lectinology” [56] se puede decir que Stillmark en 1888 identificó la primera lectinas, al describir el fenómeno de hemaglutinación producida por extractos de semillas de ricino (*Ricinus Communis*). Aisló el componente tóxico de carácter proteico responsable de dicha aglutinación de los eritrocitos, y lo denominó Ricina. James B. Sumner en 1919 [60] aisló la Concanavalina A, a partir de *Canavalia Ensiformes*. Esta fue la primera lectina vegetal

analizada y caracterizada mediante difracción de Rayos X. Con A actúa asociándose específicamente a glucosa y a manosa [61]. En 1936, junto con Howell [62], describió la capacidad de aglutinar tanto eritrocitos como células de levaduras. Esta aglutinación se inhibía por la adición de ciertos azúcares, lo que demostró, por primera vez, la especificidad de esta lectina por Glc y Man.

Morgan y Watkins, en los años 50, utilizaron la especificidad hemaglutinante del tipo de sangre para demostrar que los tipos de sangre con inmunodeterminantes (A) se unen a α -GlcNAc y los que tienen el determinante (O) se unen a α -L-Fuc. Esta fue la primera demostración de que los carbohidratos de la superficie pueden servir como soportes de información biológica [63].

Desde entonces, y según se han ido descubriendo sus propiedades en la aglutinación de eritrocitos y de otras células como linfocitos, espermatozoides, plaquetas y bacterias, inducción de mitosis en linfocitos, efectos citotóxicos sobre linfocitos, aglutinación de virus, etc. [57], las lectinas han sido objeto de interés y herramientas esenciales en muchos estudios en muy diversos campos (histología, inmunología, cáncer, infecciones por patógenos, diseño de fármacos etc.) [56, 58, 64-65].

Aunque el interés por las primeras lectinas identificadas provenía de su capacidad para aglutinar células y, por tanto, las lectinas deberían poseer por lo menos dos sitios de reconocimiento de carbohidratos, en la actualidad el término lectina se ha generalizado y se aplica a proteínas con al menos un sitio de reconocimiento. En cualquier caso, la multivalencia es una propiedad muy generalizada en las lectinas. En algunos casos, una lectina en estado monómero puede tener uno o más sitios de unión de carbohidratos. En otros casos, los monómeros de las lectinas con un solo sitio se asocian como dímeros, trímeros o tetrámeros, etc., creando un sistema polivalente que garantiza una alta especificidad para interacciones con ciertos tipos de superficies celulares. Estas presentan igualmente de una forma polivalente los correspondientes carbohidratos ligandos, aprovechando de esta manera el carácter aditivo de las interacciones múltiples [66].

En general, este grupo de proteínas de unión a carbohidratos actúan como mediadores centrales de la transferencia de información en los sistemas biológicos. Realizan sus funciones mediante la unión a glicoproteínas, glicolípidos y oligosacáridos, y presentan preferentemente selectividad para los residuos de carbohidratos terminales y/o

subterminales. Debido al papel fundamental que desempeñan las lectinas, están omnipresentes en muchos procesos vitales, y han sido identificadas en todos los seres vivos, desde microorganismos unicelulares hasta animales y plantas.

Actualmente, una gran variedad de lectinas se pueden obtener en cantidades significativas, bien por extracción directa desde la fuente natural o bien por expresión en bacterias o cultivos celulares. Esta disponibilidad hace que las lectinas, especialmente las de origen vegetal, sean un modelo muy asequible para el estudio de interacciones carbohidrato-proteína y, al mismo tiempo, una herramienta muy poderosa para la identificación de las distintas estructuras mono-, oligo- o polisacáridicas (libres o formando parte de glicoconjugados como glicoproteínas y glicolípidos) presentes en una gran variedad de sistemas biológicos.

Estas características de las lectinas las han convertido en herramientas importantes en diversas áreas de investigación en histoquímica [65], análisis estructural de cadenas de oligosacáridos, detección de la alteración de la glicosilación, cuantificación y purificación por afinidad de los glicoconjugados, estudios citofluométricos de las células y separación celular [67]. Algunas lectinas son mitogénicas, mientras que otras son citotóxicas y participan en multitud de procesos de reconocimiento célula-célula. Por ejemplo, puede mencionarse la interacción microorganismo-célula, tanto de carácter patógeno (p. e. toxina de cólera) [52], como simbiótico (reconocimiento de factores de nodulación) [68]. Las lectinas también son vehículos para la administración de fármacos [69]. Además, las lectinas de plantas ampliamente distribuidas en los alimentos, desempeñan un papel importante en nutrición y salud [65, 69-71].

2.1 CLASIFICACIÓN DE LAS LECTINAS

Se han identificado grandes variedades de lectinas, pero no existe un criterio absolutamente general para su clasificación. Dependiendo del contexto de estudio se utilizan distintos esquemas de clasificación, por ejemplo, según su origen (lectinas animales, vegetales, bacterianas, etc.), según su secuencia de DNA y relación evolutiva, según su funcionalidad y especificidad frente a carbohidratos, según su plegamiento y homología estructural, o siguiendo otros criterios como los tipos de estructuras cuaternarias o supramoleculares.

Sharon [57] agrupó a las lectinas en cinco clases, según su afinidad máxima para Man, Gal o GalNAc, GlcNAc, Fuc o ácido siálico (NeuAc, N-Acetilneuramínico). Estos monosacáridos y sus derivados son las unidades típicas de la estructura de oligosacáridos en las superficies celulares. Sin embargo, esta clasificación no es suficiente para abarcar todos los tipos de interacciones carbohidrato-lectina. Por un lado, hay otros tipos de monosacáridos que, aunque con menos frecuencia, también se han encontrado en procesos específicos de interacción carbohidrato-lectina y, por otro lado, hay lectinas cuya especificidad se expresa frente a estructuras oligosacáridicas más complejas y no se pueden definir frente a un único monosacárido [72].

Con la secuenciación de un número cada vez mayor de genomas, el número de posibles proteínas identificadas como lectinas está en continua expansión. Sin embargo, su caracterización funcional y estructural crece a un ritmo mucho más lento y ninguno de los criterios de clasificación es completamente satisfactorio por sí solo. Varki y col. han presentado una clasificación basada en una combinación de similitudes estructurales y/o de secuencia y relaciones evolutivas[73].

Imberty y col. mantienen una base de datos estructural de lectinas clasificadas primeramente en función de su origen (Figura 1.10) y seguidamente agrupadas en familias manteniendo una nomenclatura y criterios de clasificación semejantes a los que fueron usadas por Varki y col. Esta base de datos comprende, hoy en día, más de 1100 estructuras tridimensionales de lectinas en estado libre y enlazado, algunas de los cuales están representados por varias estructuras obtenidas con diferentes ligandos. La base de datos incluye 56 familias estructurales diferentes, 8 familias virales, 6 de plantas y 9 de hongos y levaduras, 15 de bacterias, 19 de animales y una de algas (Figura 1.10).

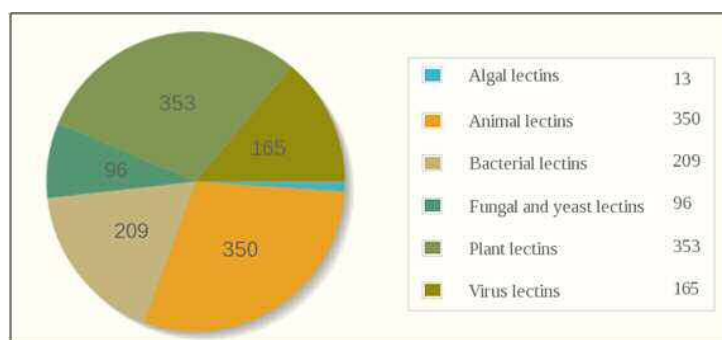


Figura 1.10 Numero de lectinas identificadas, caracterizada y recogidas en grupos según su origen en la base de datos de Imberty y col (<http://lectin3d.cermav.cnrs.fr/search.php>).

El Consorcio de Glicómica funcional (CFG, *Consortium of Functional Glycomics*, <http://www.functionalglycomics.org>), sin tratar de presentar una clasificación exhaustiva, agrupa las lectinas del consorcio en 9 familias de *glycan binding proteins* (GBPs), diferenciadas esencialmente por parámetros funcionales.

Desde el punto de vista de su estructura cuaternaria, Sharon y Lis clasificaron las lectinas en tres categorías según su topología macro- o supra macromolecular [56, 58, 64-65]:

Lectinas simples. Estas lectinas se componen de varios monómeros formando homo- o heterooligómeros, el peso molecular de cada monómero en general, no supera los 40 kDa. Cada monómero contiene un dominio con el sitio de unión a carbohidratos y, en algunos casos, también contiene otro dominio con distinta actividad. En esta categoría se incluyen la gran mayoría de las lectinas de plantas, lectinas bacterianas solubles y algunas lectinas animales como las galectinas, específicas de galactosa (Figura 1.11 a).

Lectinas en mosaico. Este grupo incluye una gran variedad de proteínas, muchas de las lectinas de origen animal, pero también de otras fuentes (virus, bacterias, plantas, etc.). Pueden presentarse en forma monomérica o multimérica (Figura 1.11 b), pero los monómeros son cadenas polipeptídicas complejas que incluyen varios tipos de módulos o dominios con distintas funcionalidades. Solo uno de estos tipos de módulo es el responsable del sitio de unión a carbohidratos, pero puede aparecer en forma pseudo-polimérica por repeticiones múltiples en la misma cadena polipeptídica. Este es el caso de algunas C-lectinas (Figura 1.11 c).

Ensamblajes macromoleculares. Las lectinas de este grupo se encuentran con frecuencia en el reino de las bacterias, donde forman estructuras filamentosas de 3 a 7 nm de diámetro y hasta 100 nm de longitud, llamadas fimbrias o pilis. La mayor parte de este filamento se forma por la polimerización de una unidad dominante, que desempeña un papel estructural. Sólo un tipo de unidades, generalmente minoritaria, tiene el sitio de unión para los carbohidratos y, por lo tanto, es responsable de la capacidad de adhesión de las fimbrias a otras células (Figura 1.11 d). Esas lectinas son importantes factores de patogenicidad.

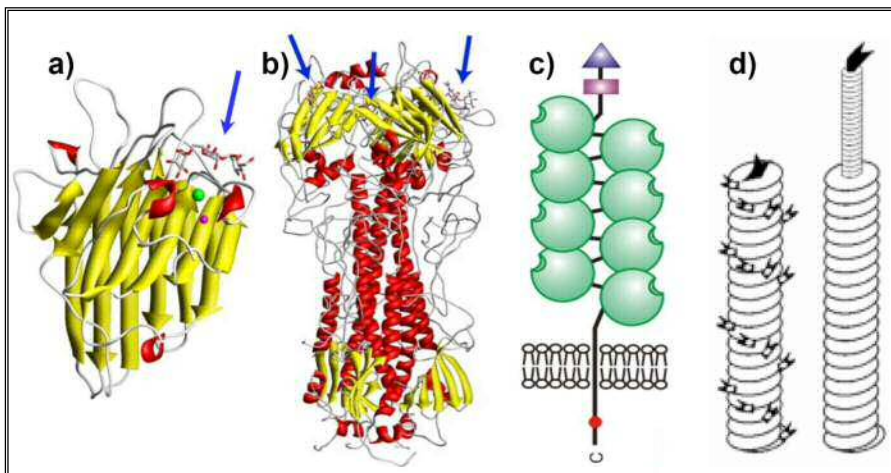


Figura 1.11 Estructuras tridimensionales de **a)** Lectina simple, monómero de Concanavalina A de *Canavalia ensiformis* en complejo con el trimanosido (código PDB 1CVN), **b)** Lectina mosaico, trímero de la hemaglutinina del virus de influenza en complejo con el ácido siálico (código PDB 1HGI), **c)** Esquema de lectina mosaico, receptor de manosa de macrófagos con hasta 8 repeticiones del *Carbohydrate recognition domain* [74] y **d)** Esquema de ensamblajes macromoleculares, ejemplos de fimbria bacterianas, la subunidad marcada en negro corresponde a la subunidad adhesina de unión a carbohidrato [64]. Las flechas azules señalan al azúcar sustrato.

2.2 LOS SITIOS DE UNIÓN DE LECTINAS

Los sitios de unión de las lectinas están constituidos generalmente por una cavidad, poco profunda, en la superficie de la proteína, preorganizada para acoger al ligando de manera que su geometría no varíe mucho después de su unión. Las lectinas interactúan con los carbohidratos mediante una red de enlaces de hidrógeno y de interacciones hidrófobas y de van der Waals. En algunas familias de lectinas, también participan cationes en el proceso de unión a carbohidratos o se dan interacciones de tipo iónico, en el caso de que los ligandos sean azúcares cargados. Además de estas generalidades, los sitios de unión a carbohidratos tienen características propias según sea la familia estructural de lectinas y su ligando específico, de modo que no es posible establecer un modelo general del sitio de unión en lectinas.

En la figura 1.12 se presentan, a modo de ejemplo, la estructura de dos complejos lectina-carbohidrato (galectinas, lectinas animales que unen galactosa y C-lectinas, con dependencia de Ca^{2+}). El primer caso corresponde a la galectina 8 unida a $\alpha(2-3)$ sialilactosa [75] (Figura 1.12 a), donde la unión proteína-azúcar está basada en los tres tipos de las interacciones descritos en el apartado 1.xx: enlaces de hidrógeno (las

argininas 45 y 69 establecen enlaces de hidrógeno con el carbohidrato), apilamiento CH/ π (el triptófano 86 estabiliza la cara apolar del segundo anillo del azúcar) e interacciones polares (el carboxilato del ácido siálico se une con el resto guanidinio de la arginina 59). El segundo caso corresponde al *sialyl-Lewis X* unido a la E-selectina humana (Figura 1.12 b). Aquí se observa un catión Ca^{2+} que está coordinado con dos hidroxilos del azúcar [76].

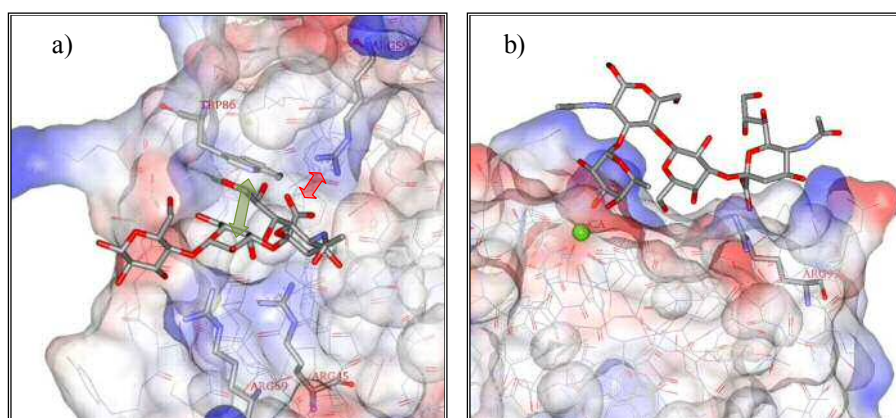


Figura 1.12 Imagen de **a)** Sitio de reconocimiento de la galectina-8 en complejo con la sialillactosa (Código PDB 3AP7), **b)** Sitio de unión de la E-selectina (C-lectina) humana en complejo con el *Sialyl Lewis X* (PDB 1G1T). El calcio está presentado por una esfera de color verde.

Desde el punto de vista termodinámico, en este tipo de sitios de unión superficiales no se suelen establecer interacciones de gran afinidad con las unidades monosacáridicas (entre 10^2 - 10^4 M^{-1}). En el caso de ligandos más complejos, como oligosacáridos, se pueden establecer más contactos con la proteína aumentando así la afinidad (Figura 1.12). Adicionalmente, la baja afinidad por monosacáridos es obviada mediante fenómenos de multivalencia. En algunos casos, la lectina presenta múltiples sitios de unión por repeticiones intramoleculares de varios dominios de unión en la misma cadena polipeptídica o por formación de estructuras multiméricas con varias cadenas de lectinas. Por otra parte, el ligando oligosacáridico puede presentarse en forma de glicoconjugados multivalentes (por ejemplo, como estructuras oligo y polisacáridicas ramificadas (Figura 1.4) o como glicoproteínas modificadas por múltiples cadenas oligosacáridicas o glicolípidos en membranas) [77].

2.2.1 Las lectinas que reconocen galactosa

Es interesante resaltar que, desde el punto de vista químico, la galactosa tiene una naturaleza inherentemente distinta a la de la glucosa, manosa y fructosa en el contexto de su biosíntesis y de su posible origen evolutivo. Los últimos tres monosacáridos se pueden interconvertir abióticamente mediante la transformación aldosa-cetosa (*Lobry de Bruyn-van Ekenstein transformation*) [78]. Por otro lado, la glucosa y la manosa son materiales de partida en la biosíntesis de otros diversos azúcares, pero no así la galactosa [79]. Estas y otras observaciones llevaron a Hirabayashi a proponer una hipótesis sobre el origen de hexosas elementales en la que la galactosa es definida como un sacárido “recién llegado” en relación a la glucosa y manosa [79-80]. En el contexto de las estructuras oligosacáridicas más complejas, está bien establecido que en la biosíntesis de cadenas de N-glicano, la galactosa se incorpora en las etapas terminales después de la eliminación de los residuos de glucosilo y manosilo desde el precursor común $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ [81-82]. Este carácter terminal hace que, en muchos casos, la galactosa quede presentada en las zonas más expuestas de las células, a menos que se enmascaren por los ácidos siálicos. De este modo, pueden ser fácilmente reconocibles o accesibles o incluso utilizadas como puntos de anclaje por diversas moléculas/receptores expresados por otras células con relación homofílica (células del mismo tipo) o heterofílicas (células distintas, de otros tejidos u otros organismos). En este contexto, es interesante resaltar la gran variedad de lectinas que reconocen galactosa en las más diversas especies. Una posible explicación para tal variedad reside en el hecho de que parece que la galactosa ha sido seleccionada evolutivamente como un carbohidrato importante para el reconocimiento molecular y, muy especialmente, en los organismos superiores. Dentro de las familias de lectinas que interaccionan con galactosa encontradas en los vertebrados, el grupo de las galectinas (anteriormente denominadas S-lectins) representa una amplia familia con más de 15 tipos descritos en animales. Se unen específicamente a la galactosa y a oligosacáridos relacionados (lactosa (Gal1-4bGlc), N-acetil-lactosamina (Gal1-4bGlcNAc), etc.) [83]. Se ha caracterizado la estructura tridimensional de un gran número de galectinas formando complejo con sus ligandos (hasta 160 estructuras relacionadas con galectinas están en el RCSB) (en la figura 1.12 a se muestra la estructura de la galectina 8). En 1998 se describió la primera estructura 3D de una galectina humana (la galactina-7 (hgal-7)) en su forma nativa y asociada a galactosa, galactosamina, lactosa, y N-acetil-lactosamina

[36]. En la familia de las C-lectinas también hay varios representantes con especificidad hacia galactósidos. En los invertebrados y las plantas también se encuentran otras familias de lectinas que reconocen galactósidos, especialmente las que tienen en común la presencia de dominios estructurales "*beta-trébol*" en su estructura y cuyo representante más característico es la ricina. Existen estructuras 3D de lectinas de tipo ricina como la lectina III (CEL-III) de *Cucumaria echinata* [84] y la lectina de *Lumbricus terrestris* [85]. En el reino vegetal, dentro de la familia de las moráceas, se encuentran lectinas que se unen a galactosa, como la jacalina extraída de las semillas de *Artocarpus integrifolia* [86]. La familia de Amaranthaceae contiene la lectina ACA *Amaranthus caudatus*, que se cristalizó con la β Gal1-3GalNAc [87]. La familia euphorbiaceae incluye lectinas que presentan una alta toxicidad. Un ejemplo de esta familia es la propia ricina o su variante tetramérica ABAB "*Ricinus communis agglutinin*" (RCA), menos tóxica, ambas aisladas de las semillas de *Ricinus communis*. La ricina es un heterodímero AB, clasificada también dentro de las proteínas inactivadoras de ribosomas tipo 2 (RIP-II), en las que la cadena B de 262 aminoácidos constituye el dominio lectina que reconoce la galactosa [88]. En los microorganismos, la familia de lectinas solubles incluye proteínas citoplásmicas como PA-IL de *Pseudomonas aeruginosa* específica para α -D-galactosa. PA-IL muestra una estructura tetramérica en la que uno o dos iones calcio están implicados en el sitio de reconocimiento al carbohidrato [89].

Las RIP de tipo II y la Viscumina:

Dentro del gran grupo de lectinas de plantas, una serie de lectinas pertenece al grupo denominado RIP-II (proteínas inactivadoras de ribosomas de tipo 2). Esta actividad inhibidora de la síntesis de proteínas les confiere un carácter tóxico y se les asigna una función defensiva en plantas [90]. El miembro más destacado de este grupo es la ricina, proteína aislada de la semilla de *Ricinus Communis* conocida desde hace más de cien años como proteína hemaglutinante y altamente tóxica. Además de la ricina y las lectinas de muérdago (ML "*Mistletoe Lectin*" o VAA "*Viscum album agglutinin*"), encontramos lectinas similares en otras plantas como, por ejemplo, la volkensina (*Adenia volkensii*, planta Kilyambiti, Somalia), abrina (*Abrus precatorius*, árbol de rosario, Perú), ebulina II (*Sambucus ebulus*, sauco menor) y otras [58, 67, 70, 72, 91-92].

Las RIPs son una familia bien definida de proteínas que pueden inhibir catalíticamente la síntesis de proteínas en células eucariotas, dañando el ribosoma [90]. Se dividen en dos grupos: RIPs de tipo I, que representan proteínas de una sola cadena sencilla que no se une a los carbohidratos y RIPs de tipo II, que engloban proteínas de 2 cadenas peptídicas: A (cadena activa) y B (cadena de unión a carbohidratos) unidas por un puente disulfuro. La cadena A posee actividad RNA N-glucosidasa y es la responsable de su actividad inactivadora de los ribosomas [93]. La cadena B es la que interviene en la primera parte del proceso; es decir, en la localización y anclaje a las células diana. Se une específicamente a los glicoconjugados expuestos en la superficie celular facilitando la entrada al interior de la célula de la cadena A, que inactiva los ribosomas [94-95]. Es en este punto donde la capacidad de reconocimiento de galactósidos cobra especial relevancia. De hecho, se ha propuesto la aplicación de estas proteínas como "*magic bullet*" en el tratamiento del cáncer capaces de identificar y anclarse a las células diana gracias a su carácter lectina y destruirla gracias a su capacidad inactivadora de ribosomas.

2.2.2 Las lectinas que reconocen N-acetilglucosamina

La GlcNAc es un carbohidrato que se polimeriza en largas cadenas, a través de enlaces β (1-4), formando la quitina, que se encuentra en los exoesqueletos de insectos. También se encuentra en la capa de péptido-glicano que protege a las células del mundo de las bacterias y en la pared celular de ciertas algas marinas. Encontramos también la N-acetilglucosamina con el ácido glucurónico en el hialuronato, un polisacárido que sirve como amortiguador y lubricante. Las lectinas que se unen a este grupo de carbohidratos se pueden encontrar en todos los seres vivos (virus, bacteria, plantas, animales etc.). A modo de ejemplo, en microorganismos como la ameba patogénica *Entamoeba histolytica*, se ha descrito una lectina que reconoce GlcNAc [96]. En invertebrados, la tachylectine 2 [97] y la tachylectine 5A, una proteína de tipo fibrinógeno [98] se asocian a GlcNAc. En plantas se ha descrito un gran número de lectinas que unen GlcNAc. En la tabla 1.1 se mencionan algunas de las lectinas de origen vegetal no relacionadas secuencialmente y con distintas estructuras.

Tabla 1.1 Ejemplos de algunas lectinas vegetales con especificidad de unión a GlcNAc.

Lectina	Origen / familia	Carbohidrato inhibidor
LEA	<i>Lycopersicum esculentum</i> / Solanáceas	D-GlcNAc
PWL	<i>Phytolacca americana</i> / fitolacáceas	D-GlcNAc
WGA	<i>Triticum vulgare</i> (germen de trigo) / Poáceas	D-GlcNAc
STA	<i>Solanum tuberosum</i> / Solanáceas	D-GlcNAc

En la base de datos CAZY (*Carbohydrate Active enZymes*) se pueden encontrar varias familias de los denominados Carbohydrate Binding Modules, CBM, para los que se ha identificado capacidad de reconocimiento de quitina o estructuras relacionadas, entre los que se pueden mencionar el CBM 18 (<http://www.cazy.org/CBM18.html>). Este módulo incluye más de 800 secuencias del denominado dominio heveína, caracterizado por su potencial de unión a la quitina y presente en muchas proteínas de defensa de plantas [99-102] o el CBM 50 (<http://www.cazy.org/CBM50.html>), que incluye los dominios LysM, de unos 50 aminoácidos. Estos dominios se han asociado a funciones de unión y reconocimiento de factores de nodulación del tipo lipoquitooligosacáridos [103-106].

Algunas lectinas de origen vegetal, denominadas dominios heveína, se unen reversiblemente a la quitina (oligómeros de GlcNAc), el componente principal del exoesqueleto de insectos y crustáceos y que está también presente en la pared celular de numerosos microorganismos. Muchos de los dominios de heveína se han asociado con las funciones de defensa antimicrobiana de las plantas. Sus propiedades estructurales intrínsecas, así como su disponibilidad, han permitido su utilización como modelos de estudio para investigar el mecanismo, el origen y las características estructurales involucradas en la formación de los complejos proteína-carbohidrato. De esta manera, se ha determinado la estructura molecular de distintos dominios de heveína, bien libres o bien unidos a carbohidratos, mediante RMN [33, 107-108]

3 INTRODUCCIÓN AL ESTUDIO DE INTERACCIONES PROTEÍNA-CARBOHIDRATO MEDIANTE RMN

Es bien conocido que, mediante el uso de técnicas modernas de espectroscopia de RMN, es posible investigar, con cierto detalle, los eventos de interacción en disolución, en condiciones próximas al medio fisiológico. Las técnicas de RMN se pueden emplear

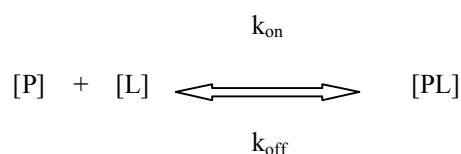
tanto para estudios de procesos de interacción proteína-ligando como proteína-proteína. Estos métodos, generalmente asistidos por protocolos computacionales, permiten elucidar las características estructurales de las entidades de unión, así como las características cinéticas y termodinámicas del equilibrio entre el estado libre y el estado asociado.

Las interacciones proteína-carbohidrato están involucradas en una gran variedad de procesos biológicos, por lo que el conocimiento de la estructura y dinámica de estos complejos es de vital importancia para comprender los fenómenos correspondientes. Desde el punto de vista de la RMN, todos los parámetros espectroscópicos, como son los desplazamientos químicos, constantes escalares, difusión, velocidades de relajación, NOE, etc., se afectan en mayor o menor medida, dependiendo de la afinidad y de la velocidad del intercambio entre los estados libre y asociado, al producirse la asociación entre ligando y receptor.

Existen dos enfoques basados en RMN que son complementarios para la identificación y estudio de estas interacciones carbohidrato-proteína:

- Siguiendo las variaciones de los parámetros de RMN de las señales del receptor al interaccionar con el ligando (carbohidrato).
- Siguiendo los efectos producidos en los parámetros de RMN del ligando al pasar del estado libre a asociado.

La unión de los ligandos (pequeños) a las proteínas receptoras (de gran tamaño) sigue una reacción de asociación biomolecular con cinética de segundo orden.



Esta ecuación representa un equilibrio dinámico en el que hay tres especies: el receptor P libre, el ligando L libre, y el complejo receptor-ligando PL. La constante de velocidad k_{off} es inversamente proporcional a la vida media del complejo receptor-ligando. La constante biomolecular k_{on} mide la probabilidad de un encuentro productivo entre el receptor y el ligando libre. Se supone que está controlada por la difusión. Por lo tanto, k_{on} puede variar entre 10^6 y $10^{10} \text{ M}^{-1} \text{ s}^{-1}$. La afinidad de unión puede ser descrita por la constante de disociación K_d , que depende de la temperatura del equilibrio.

$$K_d = [P] * [L] / [PL] = k_{off} / k_{on}$$

K_d tiene las unidades de concentración. Por lo tanto, un valor de K_d en el intervalo de mM implica una relación de aproximadamente 1:1000 de libre a estados ligados en una mezcla equimolar de P y L. Una K_d en el intervalo de μM implica una relación de aproximadamente 1:1.000.000 de estos estados; es decir, un complejo mucho más estable, con menos estados “libres” en los especies presentes. k_{on} puede variar extensamente, entre un rango de $10^4 - 10^{11}$, siendo mayor si el proceso de unión necesita reordenamientos conformacionales del ligando o del receptor. Diversos métodos de RMN se pueden utilizar para la determinación de la K_d de un complejo proteína-ligando, como se ha revisado por Fielding [109].

En el equilibrio de dos estados dado por la ecuación de arriba, las moléculas de ligando y receptor existen en estado libre (P, L) o enlazado (PL). En el estado libre, tanto el receptor como el ligando conservan sus parámetros intrínsecos de RMN (por ejemplo, los desplazamientos químicos, velocidad de relajación, coeficiente de difusión traslacional). En el estado enlazado, la afinidad mutua de la unión del ligando al receptor conduce a un proceso de intercambio, que alterna las dos moléculas entre sus estados libres y asociados. El equilibrio que adopten las poblaciones del estado libre y enlazado ($[P]$, $[L]$, $[PL]$) es consecuencia de la ecuación de arriba. Bajo estas condiciones, el ligando transitoriamente adopta los parámetros de RMN características del receptor, típicamente mucho mas grande.

En esta parte de la memoria sólo hemos abordado métodos de RMN basados en la observación del ligando; es decir, vamos a centrarnos en la detección de la existencia de interacción siguiendo las señales del ligando.

3.1 MÉTODOS DE RMN BASADOS EN LA OBSERVACIÓN DEL LIGANDO

Los métodos basados en la observación de las señales del ligando se apoyan en el hecho de que es posible distinguir entre los estados libre y asociado desde el punto de vista de la RMN, gracias a las diferencias de la velocidad rotacional y traslacional de una molécula pequeña al pasar del estado libre al asociado. Las aproximaciones basadas en la observación del ligando se basan en la transferencia, mediada por intercambio químico de la información del estado asociado al estado libre. Este requisito hace que

estos métodos sean especialmente útiles para la identificación de ligandos con afinidades entre débiles y moderadas (intercambio relativamente rápido, con k_{off} mayor de $2\text{-}3\text{ s}^{-1}$). Generalmente, los ligandos son moléculas con un peso molecular inferior a 1 kDa y se caracterizan por tener tiempos de correlación cortos, lo que se traduce en velocidades de relajación lentas y coeficientes de difusión translacional relativamente altos. Sin embargo, cuando se asocian a una macromolécula, sus propiedades de RMN cambian y adquieren las propiedades de ésta.

La ventaja de este método radica en que se evita la necesidad de producir grandes cantidades del receptor marcado isotópicamente. Además, tampoco es necesaria la asignación de la proteína, ni imprescindible el conocimiento de su estructura tridimensional. La ausencia de estos requisitos hace que estos métodos sean más rápidos y de mayor aplicabilidad al estudio de los procesos de formación de los complejos que los basados en la observación del receptor. Obviamente, la información relativa al receptor que se puede obtener con estos métodos es más limitada que en los anteriores, aunque, en casos favorables, pueden complementarse eficazmente con protocolos de docking.

3.1.1 Diferencia de Transferencia de Saturación (STD)

Esta técnica no tiene límite superior del tamaño de la molécula del receptor y se puede aplicar para detectar cualquier interacción ligando-receptor, siempre que exista la posibilidad de irradiar selectivamente las señales de RMN del receptor. Experimentalmente, para las proteínas, la irradiación selectiva se realiza típicamente en la región aromática (si los ligandos son carbohidratos) o en la región alifática de campo alto (ca -2 ppm), donde hay una densidad relativamente alta de señales de proteína y no hay resonancias de ligando [110-111]. La saturación se mantiene durante un determinado periodo (de centenares de milisegundos a unos pocos segundos) y en este intervalo, la magnetización se transfiere a través del espacio a todos los protones de la proteína. Si un ligando interacciona con la proteína, sus protones también sentirán la transferencia desde los núcleos del polipéptido y, por tanto, se afectarán sus intensidades (Figura 1.13). Si el experimento STD se realiza con una biblioteca de moléculas pequeñas, solamente se afectarán las señales de RMN de aquellas moléculas que interaccionen con el receptor. Desde el punto de vista experimental, el experimento se realiza en el modo de diferencia. Es decir, al espectro resultante de la saturación

mencionada antes se le resta el obtenido mediante saturación en una región de espectro sin señales (ca. 100 ppm). El espectro diferencia sólo muestra las señales de los ligandos que interaccionan con el receptor. Además, dependiendo del tiempo de saturación y de la velocidad de disociación del complejo, los protones del ligando que mostrarán mayor transferencia de saturación serán aquellos que estén más cerca de los protones del receptor, lo que puede permitir deducir el epítipo del ligando.

Este método se ha utilizado para determinar afinidades absolutas y relativas mediante experimentos de competición y de titulación. Pero quizás, el aspecto más elegante del método STD es que también permite conocer la región del ligando que está en contacto más íntimo con el receptor (epítipo) [112].

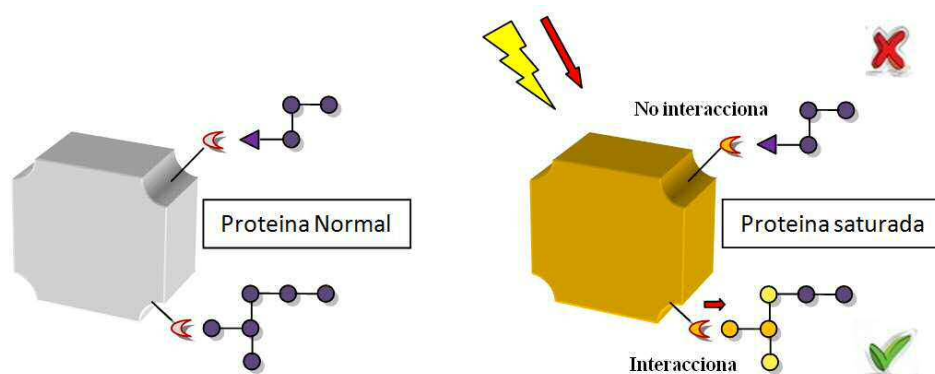


Figura 1.13 Representación de un experimento STD. Cuando el ligando interacciona con la proteína saturada, la saturación se transfiere al ligando. Los protones del ligando que están en contacto con el sitio de unión se afectan antes y más fuertemente por la saturación de los protones de la proteína.

El experimento STD presenta ciertas características que merecen ser resaltadas. Primero, el STD es ideal para el estudio de receptores de gran tamaño (> 30 kDa) que son los que generalmente se encuentran durante el proceso de descubrimiento de fármacos. Las macromoléculas con grandes pesos moleculares poseen tiempos de correlación rotacional grandes que favorecen la difusión de espín, y consecuentemente, la transferencia de saturación entre el receptor y el ligando es muy eficaz. Segundo, los experimentos STD no necesitan de altas concentraciones de receptor (~ 1 - 50 μ M). Si asumimos un intercambio lo suficientemente rápido del ligando, la población de ligandos saturados crece durante el tren de pulsos de radiofrecuencias sostenido, debido

a sus bajas velocidades de relajación en el estado libre. A través de este mecanismo, una pequeña cantidad de receptor puede lograr un gran efecto de amplificación del ligando saturado. Una tercera ventaja de este método es que sólo se observan las señales del ligando unido. No hay que hacer correcciones basadas en la contribución del estado libre que podrían complicar la interpretación. Esto es particularmente útil cuando se emplea un diseño experimental donde existe una proporción elevada ligando: receptor. Se estima que el rango de aplicación de esta técnica va desde constantes de disociación $10^{-8} < K_d < 10^{-2}$.

3.1.2 Experimentos de medidas de tiempo de relajación T_1

Las variaciones de los tiempos de relajación longitudinales T_1 o de las inversas de las velocidades de relajación, R_1 ($1/T_1$), utilizando sus versiones selectivas ($T_{1,s}$) pueden también utilizarse para detectar la existencia de entidades supramoleculares en disolución [113]. De hecho, existe una contribución muy importante del tiempo de correlación t_c al valor de $T_{1,s}$, con lo que la medida de este parámetro también proporciona información muy útil sobre la existencia o no de reconocimiento. Así, las medidas de $T_{1,s}$ se han utilizado en experimentos de cribado de moléculas mediante RMN, e incluso se han utilizado para determinar constantes de disociación en experimentos de titulación.

3.2 MÉTODOS DE RMN BASADOS EN LA OBSERVACIÓN DE LA PROTEÍNA

3.2.1 Las perturbaciones de desplazamientos químicos del receptor

El efecto producido por los diferentes ambientes químicos de los protones de una molécula se traduce en la existencia de distintos desplazamientos químicos. Obviamente, cuando existe asociación ligando-receptor, se generan variaciones en los desplazamientos químicos de ambas especies; específicamente, en los núcleos que se encuentran en el sitio de unión y cercanos a este. La estimación de estas perturbaciones se puede llevar a cabo a través de técnicas de RMN mono y bidimensionales en los casos en los que se conocen los desplazamientos químicos del receptor (Figura 1.14)

El seguimiento de estas perturbaciones de desplazamiento puede permitir llevar a cabo, mediante titulación, la determinación de parámetros termodinámicos como constante de asociación, entalpía de unión, entropía y la energía libre asociada, usando una aproximación de van't Hoff. Es decir, si el intercambio entre los estados libre y asociado es rápido en escala de tiempos de los desplazamientos químicos, se puede estimar el valor de la constante de asociación. [114] [115] En estos casos, la variación del desplazamiento químico de las señales de la proteína para las distintas concentraciones molares de ligando sigue la siguiente ecuación:

$$\delta_{\text{obs}} - \delta_{\text{libre}} = (\delta_{\text{comp}} - \delta_{\text{libre}}) [\text{PL}]/P_{\text{T}}$$

PL: es la concentración del complejo proteína-ligando ; P_{T} es la concentración total de proteína ; δ_{libre} y δ_{comp} son los desplazamientos químicos de la proteína libre y del complejo respectivamente; δ_{obs} es el desplazamiento químico observable por una concentración dada de ligando.

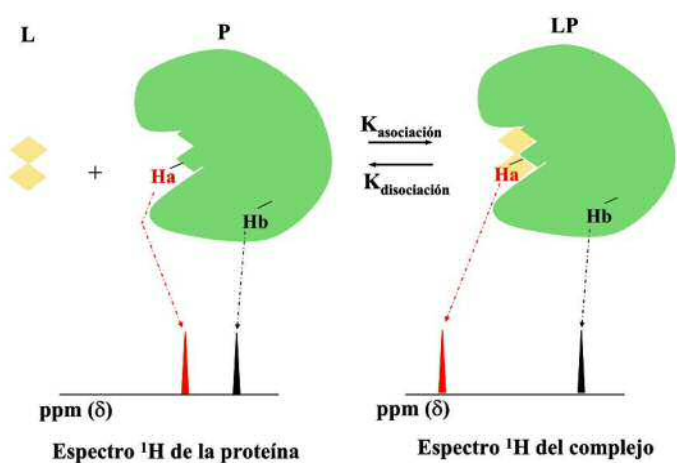


Figura 1.14 Representación de la asociación ligando-proteína mediante RMN. Se muestra la variación de desplazamiento químico para Ha (rojo), que se localiza en el sitio de reconocimiento. Sin embargo, el desplazamiento químico de Hb (negro) no se afecta.

Puesto que K asociación es directamente proporcional a la concentración de complejo e inversamente proporcional al producto de las concentraciones de la proteína y ligando, la ecuación queda definida como:

$$\delta_{\text{obs}} - \delta_{\text{libre}} = (\delta_{\text{comp}} - \delta_{\text{libre}}) K_{\text{asociación}} [P][L] / P_T$$

El experimento de perturbación de desplazamientos químicos en 2D utilizando un espectro de correlación cuántica simple heteronuclear (HSQC) ^{15}N - ^1H es muy sensible, y permite detectar perturbaciones en el ambiente de las amidas de la cadena principal, debido a la unión del ligando. Los desplazamientos químicos de las amidas también son sensibles a las perturbaciones de las cadenas-laterales conectadas, que también interaccionan con el ligando. Mediante la asignación y cálculo de los desplazamientos químicos se identifican los sitios de unión, con su correspondiente afinidad hacia el ligando. Debemos mencionar que se pueden observar cambios importantes en un residuo dado y eso puede ser interpretado de dos maneras diferentes: (1) un cambio local en la conformación proteica que altera la magnetización detectada en un núcleo en particular debido a nuevas modificaciones por los electrones que lo rodean; y (2) que el residuo está involucrado en una interacción directa con el ligando.

Comentario [j1]: llevarlo a generalidades

3 OBJETIVOS

El objetivo general de esta Tesis ha sido extender el conocimiento sobre el origen de la especificidad y estabilidad de complejos carbohidrato-proteína mediante una combinación de métodos de espectroscopía de RMN y modelado molecular. Para ello se han escogido diversas familias de compuestos proteicos (proteínas o dominios de proteínas) y de carbohidratos de distinta naturaleza química, derivados de los glicoconjugados de las paredes celulares de microorganismos, implicados en procesos de reconocimiento célula-célula. En definitiva, se ha utilizado una aproximación multidisciplinar que ha incluido el uso de técnicas diversas de RMN, de modelado molecular, métodos bioquímicos, incluyendo medidas de afinidad, y protocolos de biología molecular, incluyendo expresión y purificación de proteínas

Esta Tesis se puede dividir en dos partes bien diferenciadas recogidas en los capítulos 3 y 4:

Capítulo 3. -Modelos de interacciones de proteínas con carbohidratos derivados de N-acetilglucosamina

El objetivo de este capítulo, se ha centrado en el estudio mediante técnicas de RMN de la interacción de dominios tipo Heveína, LysM y PASTA con posibles ligandos relacionados con oligosacáridos de las paredes de microorganismos (quitoligosacáridos y fragmentos de peptidoglicano). Se han empleado métodos de RMN que permiten seguir procesos de reconocimiento molecular desde el punto de vista del ligando o desde el punto de vista del receptor. También se ha determinado la estructura 3D de una proteína.

- En el primer apartado de este capítulo, hemos planteado caracterizar, mediante técnicas de STD RMN (desde el punto de vista del ligando) la interacción de fragmentos de peptidoglicano con dominios de tipo heveína e identificar la posible participación de la parte peptídica del peptidoglicano en la interacción. Se eligió la proteína WGA extraída de las semillas del trigo *Triticum vulgaris* como modelo de lectina con dominios heveína.
- En el segundo apartado de este capítulo se ha planteado, como objetivo específico, el estudio de la interacción carbohidrato-proteína en los procesos de reconocimiento de los factores de nodulación, en los que se ha propuesto que participan dominios LysM. En este caso, se han empleado técnicas de RMN desde el punto de vista de la proteína. Para alcanzar dicho objetivo, se ha planteado la obtención de proteínas recombinantes correspondientes a estos dominios y su caracterización estructural. En una primera etapa se eligió el dominio LysM de la transglucosidasa lítica de *Escherichia coli* como un sistema modelo para el estudio de este tipo de interacciones usando como ligandos estructuras poliméricas de N-acetilglucosamina. En una siguiente etapa, se pretendía estudiar el dominio LysM de proteínas receptoras de factores de nodulación del sistema de simbiosis bacteriana en plantas. En particular, se escogió el segundo dominio LysM, presente en la parte extracelular de la proteína Lyk-3, que se había propuesto como receptor del tipo “LysM-RLK” de

los factores de nodulación de tipo LCO, en procesos de simbiosis en *Medicago truncatula* (un tipo de trébol).

- En el tercer apartado de este capítulo, se pretendió confirmar, mediante RMN y también desde el punto de vista de la proteína, la posible unión del tercer dominio PASTA de la proteína PrcK de *Bacillus subtilis* a con diferentes ligandos potenciales derivados del peptidoglicano o de antibióticos betalactámicos. Con dicho objetivo, se ha llevado a cabo la determinación de la estructura tridimensional en solución, mediante RMN, del tercer dominio PASTA de *Bacillus subtilis* para, a continuación, tratar de caracterizar los aminoácidos que participan en la posible interacción.

Capítulo 4. -Caracterización del reconocimiento diferencial de los anómeros de galactosa por la viscumina

En este capítulo, y siguiendo con nuestra línea de trabajo en este área, nos propusimos determinar, mediante la aplicación y el desarrollo de estrategias basadas en RMN desde el punto de vista del ligando, la posible selectividad de una lectina hacia cada uno de los dos anómeros α - y β - de un monosacárido, cuando estos están en equilibrio en una disolución del monosacárido. Para ello se empleó, la técnica de STD y se usaron como modelos la galactosa (como ligando) y la viscumina (como receptor).

- Nos propusimos, como primer objetivo, calcular las constantes de afinidad de los ligandos y comprobar si los mismos competían entre sí en los sitios de unión de la proteína. La caracterización de las afinidades requirió un estudio detallado de los factores que afectan a esta clase de medidas.
- Para explicar la información obtenida desde el punto de vista del ligando mediante una interpretación estructural detallada que incluyera el sitio de unión de la proteína, se dedujeron los epitopos de unión de los ligandos a partir de las intensidades de saturación STD. Esta información se sustentó con métodos de modelado molecular y de análisis con el protocolo CORCEMA.

5 REFERENCIAS BIBLIOGRÁFICAS

1. M. Taylor, and K. Drickamer., *Introduction to Glycobiology*. 2003, London: Oxford University Press.
2. C. Altona y M. Sundaralingam, *Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation*. J Am Chem Soc, 1972. **94**(23): p. 8205-8212.
3. E.E. Juaristi, *Conformational Behavior of Six-Membered Rings. Analysis, Dynamics and Stereoelectronic Effects*. 1995: VCH Publishers: New York.
4. I.T.a.T. Bleha, *Anomeric an exo-anomeric effect in carbohydrate chemistry*. Adv Carbohydr Chem Biochem, 1989. **47**: p. 45-123.
5. P.R. Schreiner, *Teaching the right reasons: lessons from the mistaken origin of the rotational barrier in ethane*. Angew Chem Int Ed Engl, 2002. **41**(19): p. 3579-3581, 3513.
6. J.P. Praly y R.U. Lemieux, *Influence of solvent on the magnitude of the anomeric effect*. Canadian journal of chemistry, 1987. **65**(1): p. 213-223.
7. E.J. Cocinero, P. Carcabal, T.D. Vaden, J.P. Simons y B.G. Davis, *Sensing the anomeric effect in a solvent-free environment*. Nature, 2011. **469**(7328): p. 76-79.
8. Y. Zhu, J. Zajicek y A.S. Serianni, *Acyclic forms of [1-(13)C]aldohexoses in aqueous solution: quantitation by (13)C NMR and deuterium isotope effects on tautomeric equilibria*. Journal of organic chemistry, 2001. **66**(19): p. 6244-6251.
9. G.G. Bouffard, K.E. Rudd y S.L. Adhya, *Dependence of lactose metabolism upon mutarotase encoded in the gal operon in Escherichia coli*. Journal of Molecular Biology, 1994. **244**(3): p. 269-278.
10. S.J. Benkovic, *Anomeric specificity of carbohydrate-utilizing enzymes*. Methods in enzymology, 1979. **63**: p. 370-379.
11. J. Salas, M. Salas, E. Vinuela y A. Sols, *Glucokinase of Rabbit Liver*. The Journal of biological chemistry, 1965. **240**: p. 1014-1018.
12. M.J. Borrok, L.L. Kiessling y K.T. Forest, *Conformational changes of glucose/galactose-binding protein illuminated by open, unliganded, and ultra-high-resolution ligand-bound structures*. Protein Sci, 2007. **16**(6): p. 1032-1041.
13. K. Gehring, P.G. Williams, J.G. Pelton, H. Morimoto y D.E. Wemmer, *Tritium NMR spectroscopy of ligand binding to maltose-binding protein*. Biochemistry, 1991. **30**(22): p. 5524-5531.
14. S. Andre, F. Javier Canada, T. Shiao, L. Largartera, T. Diercks, S. André, F.J. Cañada, M. Bergeron Brlek, K. el Biari, A. Papadopoulos, J. Ribeiro, M. Touaibia, D. Solís, M. Menéndez, J. Jiménez Barbero, R. Roy y H.-J. Gabius, *Fluorinated Carbohydrates as Lectin Ligands: Biorelevant Sensors with Capacity to Monitor Anomer Affinity in 19F-NMR-Based Inhibitor Screening*. European journal of organic chemistry, 2012. **2012**(23): p. 4354-4364.
15. H.-J. Gabius, *The Sugar Code: Fundamentals of Glycosciences*. 2009: Wiley-Blackwell.
16. R.A. Dwek, *Glycobiology: Toward Understanding the Function of Sugars*. Chem Rev, 1996. **96**(2): p. 683-720.
17. A. Varki, *Biological roles of oligosaccharides: all of the theories are correct*. Glycobiology, 1993. **3**(2): p. 97-130.
18. H.J.G. Gabius, S., *Glycosciences: Status and Perspectives*. Chapman & Hall, Londres, 1997.
19. P.I. Kitov, H. Shimizu, S.W. Homans y D.R. Bundle, *Optimization of tether length in nonglycosidically linked bivalent ligands that target sites 2 and 1 of a Shiga-like toxin*. J Am Chem Soc, 2003. **125**(11): p. 3284-3294.

20. H.J. Gabius, H.C. Siebert, S. Andre, J. Jimenez-Barbero y H. Rudiger, *Chemical biology of the sugar code*. Chembiochem, 2004. **5**(6): p. 740-764.
21. L. Mulder, B. Lefebvre, J. Cullimore y A. Imberty, *LysM domains of Medicago truncatula NFP protein involved in Nod factor perception. Glycosylation state, molecular modeling and docking of chitooligosaccharides and Nod factors*. Glycobiology, 2006. **16**(9): p. 801-809.
22. H. Kaku, Y. Nishizawa, N. Ishii-Minami, C. Akimoto-Tomiya, N. Dohmae, K. Takio, E. Minami y N. Shibuya, *Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor*. Proc Natl Acad Sci U S A, 2006. **103**(29): p. 11086-11091.
23. A. Kombrink, A. Sanchez-Vallet y B.P. Thomma, *The role of chitin detection in plant-pathogen interactions*. Microbes Infect, 2011. **13**(14-15): p. 1168-1176.
24. V. Sukhithasri, N. Nisha, L. Biswas, V. Anil Kumar y R. Biswas, *Innate immune recognition of microbial cell wall components and microbial strategies to evade such recognitions*. Microbiol Res, 2013. **168**(7): p. 396-406.
25. Y. van Kooyk y G.A. Rabinovich, *Protein-glycan interactions in the control of innate and adaptive immune responses*. Nat Immunol, 2008. **9**(6): p. 593-601.
26. L. Bhattacharyya y C.F. Brewer, *LECTIN-CARBOHYDRATE INTERACTIONS - STUDIES OF THE NATURE OF HYDROGEN-BONDING BETWEEN D-GALACTOSE AND CERTAIN D-GALACTOSE-SPECIFIC LECTINS, AND BETWEEN D-MANNOSE AND CONCAVALIN-A*. Eur. J. Biochem., 1988. **176**(1): p. 207-212.
27. A. Casnati, F. Sansone y R. Ungaro, *Peptido- and glycolixarenes: Playing with hydrogen bonds around hydrophobic cavities*. Accounts Chem. Res., 2003. **36**(4): p. 246-254.
28. E.M. Munoz, M. Lopez de la Paz, J. Jimenez-Barbero, G. Ellis, M. Perez y C. Vicent, *The relevance of carbohydrate hydrogen-bonding cooperativity effects: a cooperative 1,2-trans-diaxial diol and amido alcohol hydrogen-bonding array as an efficient carbohydrate-phosphate binding motif in nonpolar media*. Chemistry, 2002. **8**(8): p. 1908-1914.
29. V. Vicente, J. Martin, J. Jimenez-Barbero, J.L. Chiara y C. Vicent, *Hydrogen-bonding cooperativity: using an intramolecular hydrogen bond to design a carbohydrate derivative with a cooperative hydrogen-bond donor centre*. Chemistry, 2004. **10**(17): p. 4240-4251.
30. J.L. Asensio, F.J. Canada, H.C. Siebert, J. Laynez, A. Poveda, P.M. Nieto, U.M. Soedjanaamadja, H.J. Gabius y J. Jimenez-Barbero, *Structural basis for chitin recognition by defense proteins: GlcNAc residues are bound in a multivalent fashion by extended binding sites in hevein domains*. Chem Biol, 2000. **7**(7): p. 529-543.
31. F.A. Quiocho, *Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions*. Annu Rev Biochem, 1986. **55**: p. 287-315.
32. A. Nurisso, B. Blanchard, A. Audfray, L. Rydner, S. Oscarson, A. Varrot y A. Imberty, *Role of water molecules in structure and energetics of Pseudomonas aeruginosa lectin I interacting with disaccharides*. Journal of biological chemistry, 2010. **285**(26): p. 20316-20327.
33. J.L. Asensio, F.J. Canada, M. Bruix, A. Rodriguez-Romero y J. Jimenez-Barbero, *The interaction of hevein with N-acetylglucosamine-containing oligosaccharides. Solution structure of hevein complexed to chitobiose*. Eur J Biochem, 1995. **230**(2): p. 621-633.
34. J.L. Asensio, A. Arda, F.J. Canada y J. Jimenez-Barbero, *Carbohydrate-aromatic interactions*. Accounts of chemical research, 2013. **46**(4): p. 946-954.
35. M. del Carmen Fernandez-Alonso, F.J. Canada, J. Jimenez-Barbero y G. Cuevas, *Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions*. Journal of the American Chemical Society, 2005. **127**(20): p. 7379-7386.

36. D.D. Leonidas, Vatzaki, E. H., Vorum, H., Celis, J. E., Madsen, P. and Acharya, K. R. , *Structural Basis for the Recognition of Carbohydrates by Human Galectin-7*. *Biochemistry*, 1998. **37**: p. 13930-13940.
37. F.-A.M.D. Diaz M. D, Cuevas G, Canada F. J and Jimenez-Barbero J. , *Pure Appl. Chem*, 2008. **80**: p. 1827.
38. F.J.C. J. Jimenez-Barbero, G. Cuevas, J. L. Asensio, N. Aboitiz, A. Canales, M. I. Chavez, M. C. Fernandez-Alonso, A. Garcia-Herrero, S. Mari and P. Vidal, *NMR Spectroscopy and Computer Modeling of Carbohydrates: Recent Advances*. 2006. **930**.
39. A. Bernardi, D. Arosio, D. Potenza, I. Sanchez-Medina, S. Mari, F.J. Canada y J. Jimenez-Barbero, *Intramolecular carbohydrate-aromatic interactions and intermolecular van der Waals interactions enhance the molecular recognition ability of GM1 glycomimetics for cholera toxin*. *Chemistry*, 2004. **10**(18): p. 4395.
40. M.I. Chavez, C. Andreu, P. Vidal, N. Aboitiz, F. Freire, P. Groves, J.L. Asensio, G. Asensio, M. Muraki, F.J. Canada y J. Jimenez-Barbero, *On the importance of carbohydrate-aromatic interactions for the molecular recognition of oligosaccharides by proteins: NMR studies of the structure and binding affinity of AcAMP2-like peptides with non-natural naphthyl and fluoroaromatic residues*. *Chemistry*, 2005. **11**(23): p. 7060-7074.
41. S. Vandebussche, D. Diaz, M.C. Fernandez-Alonso, W. Pan, S.P. Vincent, G. Cuevas, F.J. Canada, J. Jimenez-Barbero y K. Bartik, *Aromatic-carbohydrate interactions: an NMR and computational study of model systems*. *Chemistry*, 2008. **14**(25): p. 7570-7578.
42. A.G. Santana, E. Jimenez-Moreno, A.M. Gomez, F. Corzana, C. Gonzalez, G. Jimenez-Oses, J. Jimenez-Barbero y J.L. Asensio, *A Dynamic Combinatorial Approach for the Analysis of Weak Carbohydrate/Aromatic Complexes: Dissecting Facial Selectivity in CH/ π Stacking Interactions*. *Journal of the American Chemical Society*, 2013. **135**(9): p. 3347-3350.
43. A.R. Kolatkar y W.I. Weis, *Structural basis of galactose recognition by C-type animal lectins*. *J Biol Chem*, 1996. **271**(12): p. 6679-6685.
44. C.S. Wright, 2.2 *A resolution structure analysis of two refined N-acetylneuraminyllactose--wheat germ agglutinin isolectin complexes*. *J Mol Biol*, 1990. **215**(4): p. 635-651.
45. T. Lutteke, M. Frank y C.W. von der Lieth, *Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB*. *Nucleic Acids Res*, 2005. **33**(Database issue): p. D242-246.
46. N. Sharon, *Atomic basis of protein-carbohydrate interactions: An overview*. Cambridge ed. 2006: RSC Publishing.
47. Y. Bourne, van Tilbeurgh, H. and Cambillau, C, *Protein-carbohydrate interactions*. Vol. 3. 1993: Curr. Opin. Struct. Biol.
48. D.R.a.Y. Bundle, N.M. , *Carbohydrate-protein interactions in antibodies and lectins*. Vol. 2. 1992: Curr. Opin. Struct. Biol.
49. C. Clarke, J. Gluska, A. Cooper, R.J. Woods, M.A. Nutley y G.J. Boons, *Involvement of water in carbohydrate-protein binding*. *Journal of the American Chemical Society*, 2001. **123**(49): p. 12238-12247.
50. K.A. Watson, E.P. Mitchell, L.N. Johnson, J.C. Son, C.J. Bichard, M.G. Orchard, G.W. Fleet, N.G. Oikonomakos, D.D. Leonidas, M. Kontou y et al., *Design of inhibitors of glycogen phosphorylase: a study of alpha- and beta-C-glucosides and 1-thio-beta-D-glucose compounds*. *Biochemistry*, 1994. **33**(19): p. 5745-5758.
51. E.A. Merritt, S. Sarfaty, F. van den Akker, C. L'Hoir, J.A. Martial y W.G. Hol, *Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide*. *Protein Sci*, 1994. **3**(2): p. 166-175.

52. E.A. Merritt, T.K. Sixma, K.H. Kalk, B.A. van Zanten y W.G. Hol, *Galactose-binding site in Escherichia coli heat-labile enterotoxin (LT) and cholera toxin (CT)*. Mol Microbiol, 1994. **13**(4): p. 745-753.
53. T.K. Sixma, S.E. Pronk, K.H. Kalk, B.A. van Zanten, A.M. Berghuis y W.G. Hol, *Lactose binding to heat-labile enterotoxin revealed by X-ray crystallography*. Nature, 1992. **355**(6360): p. 561-564.
54. S.M. Tschampel y R.J. Woods, *Quantifying the role of water in protein-carbohydrate interactions*. J. Phys. Chem. A, 2003. **107**(43): p. 9175-9181.
55. W.C. Boyd y E. Shapleigh, *Specific Precipitating Activity of Plant Agglutinins (Lectins)*. Science, 1954. **119**(3091): p. 419.
56. H.-J. Gabius, *The History of Lectinology*, in *The Sugar Code: Fundamentals of Glycosciences*, H.-J. Gabius, Editor. 2009, Wiley-VCH,: Wiley-VCH, Weinheim, Germany.
57. N. Sharon y H. Lis, *Lectins as cell recognition molecules*. Science, 1989. **246**(4927): p. 227-234.
58. A.a.B. Pusztai, S., *Lectins-Biomedical Perspectives*. Taylor and Francis, London ed. 1995.
59. H.-J. Gabius, S. Andre, J. Jimenez Barbero, H. Rudiger, H.-C. Siebert, S. Andr  , J. Jim  nez-Barbero y H. R  diger, *Chemical biology of the sugar code*. ChemBioChem, 2004. **5**(6): p. 740-764.
60. J. Sumner, *The globulins of the jack bean, Canavalia ensiformis*. Journal of Biological Chemistry, 1919. **37**: p. 137-142.
61. T.P. Ko, J.D. Ng y A. McPherson, *The three-dimensional structure of canavalin from jack bean (Canavalia ensiformis)*. Plant Physiol, 1993. **101**(3): p. 729-744.
62. H.S. Sumner JB, *The identification of hemagglutinin of the jack bean with concanavalin A*. Journal of Bacteriology, 1936. **32**: p. 227-237.
63. W.T. Morgan y W.M. Watkins, *The inhibition of the haemagglutinins in plant seeds by human blood group substances and simple sugars*. British journal of experimental pathology, 1953. **34**(1): p. 94-103.
64. H. Lis y N. Sharon, *Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition*. Chem Rev, 1998. **98**(2): p. 637-674.
65. S.A. Brooks, Leathem, A. J. C. & Schuhmacher, U. , *Lectin Histochemistry: a Consise Practical Handbook*. Oxford: Bios Scientific Publishers., 1997.
66. Y.C.a.L. Lee, R. T., *Carbohydrate-protein interactions: basis of glycobiology*. Acc. Chem. Res, 1995. **28**: p. 321-327.
67. M.J.D. Rhodes, *Lectin Methods and Protocols*. Humana Press ed. 161737024X, 9781617370243. Vol. 9 de Methods in Molecular Medicine. 1998. 650.
68. H.P. Spaink, *Root nodulation and infection factors produced by rhizobial bacteria*. Annu Rev Microbiol, 2000. **54**: p. 257-288.
69. E. Haltner, G. Borchard y C.M. Lehr, *Absorption enhancement by lectin-mediated endo- and transcytosis*. Methods Mol Med, 1998. **9**: p. 567-581.
70. A.a.B. Pusztai, S., *Plant Lectins*. Cambridge University Press, Cambridge ed. 1991.
71. M. Jimenez, S. Andre, H.C. Siebert, H.J. Gabius y D. Solis, *AB-type lectin (toxin/agglutinin) from mistletoe: differences in affinity of the two galactoside-binding Trp/Tyr-sites and regulation of their functionality by monomer/dimer equilibrium*. Glycobiology, 2006. **16**(10): p. 926-937.
72. E.J.M. Van Damme, Peumans, W.J., Barre, A., and Rouge, P. , *Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles*. Critical Reviews in Plant Science, 1998. **17**: p. 575-692.
73. A. Varki, M.E. Etzler, R.D. Cummings y J.D. Esko, *Discovery and Classification of Glycan-Binding Proteins*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009: Cold Spring Harbor (NY).

74. J. Arnaud, A. Audfray y A. Imberty, *Binding sugars: from natural lectins to synthetic receptors and engineered neolectins*. Chemical Society reviews, 2013. **42**(11): p. 4798-4813.
75. H. Ideo, T. Matsuzaka, T. Nonaka, A. Seko y K. Yamashita, *Galectin-8-N-domain recognition mechanism for sialylated and sulfated glycans*. J Biol Chem, 2011. **286**(13): p. 11346-11355.
76. W.S. Somers, J. Tang, G.D. Shaw y R.T. Camphausen, *Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1*. Cell, 2000. **103**(3): p. 467-479.
77. J. Lundquist y E. Toone, *The cluster glycoside effect*. Chemical reviews, 2002. **102**(2): p. 555-578.
78. J.C. Speck, Jr., *The Lobry de Bruyn-Alberda van Ekenstein transformation*. Adv Carbohydr Chem, 1958. **13**: p. 63-103.
79. J. Hirabayashi, *On the origin of elementary hexoses*. Q Rev Biol, 1996. **71**: p. 365-380.
80. J. Hirabayashi, *On the origin of glycome and saccharide recognition*. Trends in Glycoscience and Glycotechnology, 2004. **16**(88): p. 63-85.
81. K.W. Moremane, Trimble, R. B., and Herscovics, A. , *Glycosidases of the Asparagine-linked Oligosaccharide Processing Pathway*. Glycobiology, 1994. **4**: p. 113-125.
82. a.V. Trimble R. B., M. F. , *Glycoprotein oligosaccharide synthesis and processing in Yeast*. Trends Glycosci. Glycotechnol, 1995. **7**: p. 1-30.
83. H. Leffler, S. Carlsson, M. Hedlund, Y. Qian y F. Poirier, *Introduction to galectins*. Glycoconj J, 2004. **19**(7-9): p. 433-440.
84. T. Uchida, T. Yamasaki, S. Eto, H. Sugawara, G. Kurisu, A. Nakagawa, M. Kusunoki y T. Hatakeyama, *Crystal structure of the hemolytic lectin CEL-III isolated from the marine invertebrate Cucumaria echinata: implications of domain structure for its membrane pore-formation mechanism*. J Biol Chem, 2004. **279**(35): p. 37133-37141.
85. R. Suzuki, Z. Fujimoto, A. Kuno, J. Hirabayashi, K. Kasai y T. Hasegawa, *Crystallization and preliminary X-ray crystallographic studies of the C-terminal domain of galactose-binding lectin EW29 from the earthworm Lumbricus terrestris*. Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 10): p. 1895-1896.
86. R. Sankaranarayanan, K. Sekar, R. Banerjee, V. Sharma, A. Surolia y M. Vijayan, *A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a beta-prism fold*. Nat Struct Biol, 1996. **3**(7): p. 596-603.
87. T.R. Transue, A.K. Smith, H. Mo, I.J. Goldstein y M.A. Saper, *Structure of benzyl T-antigen disaccharide bound to Amaranthus caudatus agglutinin*. Nat Struct Biol, 1997. **4**(10): p. 779-783.
88. E. Rutenber, B.J. Katzin, S. Ernst, E.J. Collins, D. Mlsna, M.P. Ready y J.D. Robertus, *Crystallographic refinement of ricin to 2.5 Å*. Proteins, 1991. **10**(3): p. 240-250.
89. A. Imberty, M. wimmerova, E.P. Mitchell y N. Gilboa-Garber, *Structures of the lectins from Pseudomonas aeruginosa: insight into the molecular basis for host glycan recognition*. Microbes Infect, 2004. **6**(2): p. 221-228.
90. L. Barbieri, M.G. Battelli y F. Stirpe, *Ribosome-inactivating proteins from plants*. Biochim Biophys Acta, 1993. **1154**(3-4): p. 237-282.
91. E.J.M. Van Damme, Peumans, W.J., Pusztai, A., and Bardocz, S, *Handbook of Plant Lectins*. Wiley and Sons, Chichester, New York, Weinheim Brisbane, Singapore, Toronto, 1998: p. . 417-421.
92. T. Hajto, K. Hostanska, T. Berki, L. Palinkas, F. Boldizsar y P. Nameth, *Oncopharmacological perspectives of a plant lectin (Viscum album agglutinin-I): Overview of recent results from in vitro experiments and in vivo animal models, and their possible relevance for clinical applications*. Evid.-based Complement Altern. Med., 2005. **2**(1): p. 59-67.

93. Y. Endo, K. Mitsui, M. Motizuki y K. Tsurugi, *The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and the characteristics of the modification in 28 S ribosomal RNA caused by the toxins*. J Biol Chem, 1987. **262**(12): p. 5908-5912.
94. Y. Endo, Tsurugi, K. & Franz, H., *The site of action of the A-chain of mistletoe lectin I on eukaryotic ribosomes. The RNA N-glycosidase activity of the protein*. FEBS Lett., 1988. **231**: p. 378-380.
95. R. Mikeska, R. Wacker, R. Arni, T.P. Singh, A. Mikhailov, A. Gabdoulkhakov, W. Voelter y C. Betzel, *Mistletoe lectin I in complex with galactose and lactose reveals distinct sugar-binding properties*. Acta Crystallogr Sect F Struct Biol Cryst Commun, 2005. **61**(Pt 1): p. 17-25.
96. D. Mirelman, S. Moody-Haupt, S. Becker, Y. Nuchamowitz, R. Bracha y R. Alon, *Cell surface molecules of pathogenic and nonpathogenic Entamoeba histolytica and their relation to virulence*. Adv Exp Med Biol, 1996. **408**: p. 225-227.
97. H.G. Beisel, S. Kawabata, S. Iwanaga, R. Huber y W. Bode, *Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab Tachypleus tridentatus*. EMBO J, 1999. **18**(9): p. 2313-2322.
98. N. Kairies, H.G. Beisel, P. Fuentes-Prior, R. Tsuda, T. Muta, S. Iwanaga, W. Bode, R. Huber y S. Kawabata, *The 2.0-Å crystal structure of tachylectin 5A provides evidence for the common origin of the innate immunity and the blood coagulation systems*. Proc Natl Acad Sci U S A, 2001. **98**(24): p. 13519-13524.
99. J.L. Asensio, F.J. Cañada, H.-C. Siebert, J. Laynez, A. Poveda, P.M. Nieto, U.M. Soedjanaamadja, H.-J. Gabius y J. Jiménez-Barbero, *Structural basis for chitin recognition by defense proteins: GlcNAc residues are bound in a multivalent fashion by extended binding sites in hevein domains*. Chemistry & Biology, 2000. **7**(7): p. 529-543.
100. M.I. Chávez, M. Vila-Perelló, F.J. Cañada, D. Andreu y J. Jiménez-Barbero, *Effect of a serine-to-aspartate replacement on the recognition of chitin oligosaccharides by truncated hevein. A 3D view by using NMR*. Carbohydrate Research, 2010. **345**(10): p. 1461-1468.
101. G. Colombo, M. Meli, J. Cañada, J.L. Asensio y J. Jiménez-Barbero, *Toward the understanding of the structure and dynamics of protein-carbohydrate interactions: molecular dynamics studies of the complexes between hevein and oligosaccharidic ligands*. Carbohydrate Research, 2004. **339**(5): p. 985-994.
102. G. Colombo, M. Meli, J. Cañada, J.L. Asensio y J. Jimenez-Barbero, *A dynamic perspective on the molecular recognition of chitooligosaccharide ligands by hevein domains*. Carbohydrate Research, 2005. **340**(5): p. 1039-1049.
103. C. Gough, *Rhizobium Symbiosis: Insight into Nod Factor Receptors*. Current Biology, 2003. **13**(24): p. R973-R975.
104. F. Debellé, *Nod Factors*, in *Brenner's Encyclopedia of Genetics (Second Edition)*, S. Maloy and K. Hughes, Editors. 2013, Academic Press: San Diego. p. 74-77.
105. A. Niebel, F. Gressent, J.-J. Bono, R. Ranjeva y J. Cullimore, *Recent advances in the study of Nod factor perception and signal transduction*. Biochimie, 1999. **81**(6): p. 669-674.
106. J.V. Cullimore, R. Ranjeva y J.-J. Bono, *Perception of lipo-chitooligosaccharidic Nod factors in legumes*. Trends in Plant Science, 2001. **6**(1): p. 24-30.
107. J.F. Espinosa, J.L. Asensio, J.L. Garcia, J. Laynez, M. Bruix, C. Wright, H.C. Siebert, H.J. Gabius, F.J. Canada y J. Jimenez-Barbero, *NMR investigations of protein-carbohydrate interactions binding studies and refined three-dimensional solution structure of the complex between the B domain of wheat germ agglutinin and N,N', N''-triacetylchitotriose*. Eur J Biochem, 2000. **267**(13): p. 3965-3978.

108. A.a.J.-B. Poveda, J, *NMR studies of carbohydrate-protein interactions in solution*. Chemical Society Reviews 1998. **27**: p. 133-143.
109. L. Fielding, *NMR methods for the determination of protein-ligand dissociation constants*. Progress in Nuclear Magnetic Resonance Spectroscopy, 2007. **51**: p. 219-242.
110. R. Meinecke y B. Meyer, *Determination of the binding specificity of an integral membrane protein by saturation transfer difference NMR: RGD peptide ligands binding to integrin α IIb β 3*. J Med Chem, 2001. **44**(19): p. 3059-3065.
111. B. Claasen, M. Axmann, R. Meinecke y B. Meyer, *Direct observation of ligand binding to membrane proteins in living cells by a saturation transfer double difference (STDD) NMR spectroscopy method shows a significantly higher affinity of integrin α IIb β 3 in native platelets than in liposomes*. J Am Chem Soc, 2005. **127**(3): p. 916-919.
112. M. Mayer y B. Meyer, *Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor*. J Am Chem Soc, 2001. **123**(25): p. 6108-6117.
113. M. Delfini, R. Gianferri, V. Dubbini, C. Manetti, E. Gaggelli y G. Valensin, *^1H NMR relaxation investigation of inhibitors interacting with Torpedo californica acetylcholinesterase*. J Magn Reson, 2000. **144**(1): p. 129-133.
114. A. Poveda y J. Jimenez-Barbero, *NMR studies of carbohydrate-protein interactions in solution*. Chemical Society Reviews, 1998. **27**(2): p. 133-143.
115. J.L. Asensio, F.J. Canada, M. Bruix, A. Rodriguezromero y J. Jimenezbarbero, *The Interaction Of Hevein With N-Acetylglucosamine-Containing Oligosaccharides - Solution Structure Of Hevein Complexed To Chitobiose*. European Journal of Biochemistry, 1995. **230**(2): p. 621-633.

CAPITULO 2

MATERIALES Y MÉTODOS

Capítulo 2: Materiales y métodos

1 MATERIALES

1.1 SECUENCIAS Y PROTEINAS

- 1) La lectina del trigo WGA, fue comprada de Sigma-Aldrich (Nº Referencia 9640).
- 2) Los plásmidos clonados con la secuencia del segundo dominio LysM (Residuos 389-452) de *E.coli* (GenBank numero de acceso AAC73316), fueron amablemente cedidos por el Dr. J. Clarke, del *Centre for Protein Engineering* en Cambridge (UK).
- 3) El tercer dominio PASTA de la región extracelular de *B. subtilis* (residuos 493-559 de CE-PrkC), fue cedido por la Dra. Rita Berisio del Instituto de

Bioestructuras y Bioimágenes, del *Consiglio Nazionale delle Ricerche* (CNR) en Napoles (Italia).

- 4) La viscumina, fue purificada y cedida por el Dr. Hans-Joachim Gabius de la Universidad LFW, en Munich (Alemania).

1.2 CARBOHIDRATOS

- 1) Los oligosacáridos N,N,N-triacetil-quitotriosa y N,N,N,N,N-pentaacetil-quitopentaosa fueron adquiridos de *Toronto Research Chemical* (Nº Referencia T735000 y Nº Referencia P225000 respectivamente).
- 2) El compuesto N-Acetyl-D-glucosaminy-(β 1,4)-N-acetylmuramyl-L-alanyl-D-isoglutamine (GMDP) fue suministrado de *Calbiochem* (Nº Referencia 110150).
- 3) La resina de quitina o *chitin beads* fue comprada de *New England Biolabs* (Nº Referencia 319-93)
- 4) El peptidoglicano (PGN) de la pared de *E. coli* fue cedido por el Dr. Juan Ayala, del Centro de Biología Molecular (CBM) en Madrid (España).
- 5) El muropéptido de *B. subtilis* fue cedido por el Dr. Antonio Molinaro en Napoles (Italia).
- 6) La ampicilina en forma de sal sódica fue comprada de *Sigma-Aldrich* (Nº Referencia BCBB3794).
- 7) El tripetido N-Acetil-L-Lisina-D-Alanina-D-Alanina fue comprado de *Bachem* (Nº Referencia H-4085)
- 8) La galactosa y sus derivados α y β -Metil Galactosido fueron comprados de *Sigma Aldrich* (Nº Referencias 59-23-4 y 3396-99-4 y 1824-94-8 respectivamente).

2 MÉTODOS

2.1 TÉCNICAS DE BIOLOGÍA MOLECULAR

2.1.1 Clonaje y transformación del DNA plasmidico

La clonación consiste al proceso de aislar una secuencia de DNA de interés, insertarlo en un plásmido, generando así un DNA recombinante y obtener múltiples copias de él en un organismo.

Primero, Se debilita la membrana de las cepas bacterianas receptoras del nuevo plásmido siguiendo el protocolo de Dagert [1] que implica la utilización de CaCl_2 a 100mM. Las células competentes se transformaron luego con los plásmidos recombinantes mediante el un choque eléctrico de las células competentes por un solo pulso de 2.5 kV en cubetas de electroporación de 0.2 cm utilizando un electroporador (BioRad). Las células a continuación se incuban en medio cultivo rico en alimentos para favorecer su recuperación. Luego se seleccionan las células transformantes que han incorporado el DNA en placas que tienen un medio selectivo con presencia del antibiótico al que resiste el nuevo plásmido.

2.1.2 Electroforesis

Electroforesis de ácidos nucleicos en geles de agarosa

Se sigue el protocolo descrito por [2]. El DNA plasmidico se analizó en geles horizontales de agarosa al 0,8% (p/v), en tampón Tris /ácido acético/EDTA (0.4M/200mM/20mM), desarrolladas a 100 voltios que se corren durante 30-45 min. Finalmente, el DNA se visualiza tras la tinción con bromuro de etidio (10 mg/ml) a través de su iluminación con luz UV a 310 nm.

Electroforesis en geles de poliacrilamida

Las electroforesis se lleva a cabo en un sistema discontinua siguiendo los protocolo de laemmli [2]. La polimerización de los geles y el desarrollo de la electroforesis se llevaron a cabo en un sistema *Mini-Protean* de *BioRad* usando geles de tamaño 10 x 8 cm y de 1 mm de grosor y empleando sistemas de 10 pocillos. En cada pocillo se aplicaron muestras de extractos crudos o de purificaciones en un volumen final de 10 μl .

El gel separador se prepara con un 15% de acrilamida y bisacrilamida al 0.5% en Tris 0.38 M, pH 8.8 y SDS 0.1% (p/v). Para la polimerización de los geles se añaden disoluciones de persulfato amónico al 0.02% (p/v), como agente catalizador, y N,N,N',N'-tetrametiletilendiamina (TEMED) al 0.075%, como propagador de la reacción.

Las muestras antes de cargarlas en el gel, se disuelven en el tampón de electroforesis que contiene (Tris/HCl 80mM a pH 6.8, DTT 10mM, 2% (p/v) SDS, β -mercaptoetanol 5% (v/v), 6M urea, azul de bromofenol 0.1% (p/v) para la tinción). Se resuspenden las muestras con el tampón de carga en una relación 1: 1, y se incuban durante 10 minutos a 95°C.

La electroforesis se desarrolla a temperatura ambiente en tampón Tris/glicina/SDS (25mM/192mM/0,1%) a pH 8,3 y a 30 mA por gel hasta que el marcador llega al final del gel. En los geles se incluyeron un pocillo muestra de patrones de proteínas de bajo peso molecular usando marcadores de BioRad (Nº referencia 4110029).

Finalizada la electroforesis, las proteínas se detectan mediante tinción con azul de Coomassie G-250 (Coomassie Brilliant Blue G-250 siguiendo las instrucciones del fabricante). El exceso de reactivo se elimina mediante lavajes sucesivos de los geles en agua MiliQ.

2.1.3 Expresión y purificación del péptido recombinante

Los péptidos recombinantes se expresaron como péptidos de fusión con una extensión de 6 histidinas en el extremo N-terminal. Se realiza la prueba de expresión a pequeña escala (10 ml) seleccionando varios transformantes de *E. coli* con objeto de determinar las condiciones más adecuadas para la inducción de la expresión. Los parámetros que se analizan durante la optimización de la misma fueron: los medio de cultivo suplementados con antibiótico, momento de inducción en la fase de crecimiento bacteriano (entre 0,2 hasta 1,7), temperatura (16, 25 y 37°C), concentración de IPTG (0,1 a 1mM) y tiempo de inducción (4, 6 y 20 horas). En cada caso la producción del péptido se determinó mediante SDS-PAGE al 15%.

Establecidos los parámetros óptimos de crecimiento, se llevo a cabo la producción a gran escala.

2.1.4 Purificación del péptido recombinante

2.1.4.1 Columna de afinidad de níquel

La purificación de proteínas con una extensión de seis histidinas se llevó a cabo por cromatografía de afinidad empleando la resina Ni-ácido nitrilotriacético-agarosa (Ni-NTA-agarosa). De esta forma se dispone de un método rápido de purificación en un solo paso.

La unión de la proteína a la resina se produce a través de los seis residuos de histidina. En la figura 2.1 se observa la interacción que se produce entre los anillos imidazólicos de los residuos de histidina y los iones Ni^{2+} , así como el resto de las posiciones de coordinación de dicho ión. El ácido nitrilotriacético es el que queda unido a la resina. La capacidad de unión de la resina de Ni-NTA-agarosa es de 5-10 mg de proteína por ml. Las proteínas que contienen una extensión de seis histidinas se unen a la resina de Ni-NTA con una afinidad muy elevada ($K_d = 10^{-13}$ M a pH 8.0). Por ello, las proteínas contaminantes, que se unen inespecíficamente, pueden ser eliminadas fácilmente lavando la columna con disoluciones que contienen imidazol a concentraciones entre 10 y 50 mM. Además, la unión de proteínas con extensiones de seis histidinas no requiere una estructura determinada, por lo que no se ve afectada por la presencia de agentes desnaturalizantes como urea o cloruro de guanidinio, de detergentes no iónicos o de concentraciones elevadas de sal (0.1-1.0 M).

En líneas generales el proceso de purificación es el siguiente: la muestra a purificar se aplica en una columna de Ni-NTA-agarosa equilibrada en el tampón de carga, manteniendo el flujo entre 0.25 y 0.5 ml/min. La columna se lava con el tampón de carga con imidazol 10 mM y, a continuación, con este mismo tampón con imidazol 30 mM, hasta que la absorción a 280 nm disminuya a menos de 0.03. La proteína unida específicamente a la columna se eluye con imidazol 200 mM, recogándose fracciones de 1-1.2 ml.

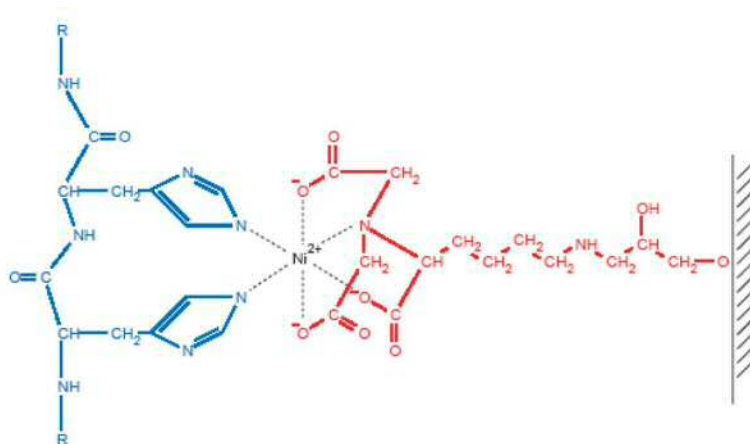


Figura 2.1 Interacción entre los anillos imidazólicos de los residuos de histidina y los iones Ni^{2+} . Los iones Ni^{2+} tienen dos de sus posiciones de coordinación ocupadas por dos átomos de nitrógeno de los anillos imidazólicos de los residuos de histidina. El resto de las posiciones están ocupadas por tres átomos de oxígeno y uno de nitrógeno del ácido nitriloacético, el cual queda unido a la resina mediante un enlace éter.

2.1.4.1 Cromatografía de filtración en gel

La cromatografía de filtración en gel es una técnica que permite separar moléculas en función de su tamaño molecular. La capacidad separadora reside fundamentalmente en el gel cuya matriz consta de un gran número de esferas porosas microscópicas. Cada gel se caracteriza por un rango de fraccionamiento que depende del tamaño de sus poros.

Las separaciones se realizaron en Columna Superdex tipo Hiload 16/60 y 26/60 de separación por filtración en gel.

2.1.5 Método de Bradford

La valoración de la concentración de proteína total se realizó mediante el método de colorimétrico de Bradford [3], utilizando una curva patrón que se realizó a partir de una solución madre de BSA (albúmina de suero bovino) de 2 mg/ml. Los resultados se expresan en $\mu\text{g}/\mu\text{L}$.

2.2 TÉCNICAS DE RESONANCIA MAGNÉTICA NUCLEAR

2.2.1 Preparación y adquisición de las muestras

Asignación del Triacetil-quitotriosa

Se preparo una disolución de la triacetil-quitotriosa libre (a 2 mM) en agua (con 10% D₂O) y se adquirieron los espectros ROESY a 400 ms [4] y TOCSY a 20 y 60 ms en un espectrómetro Bruker Avance 600MHz a 25°C.

Estudio de interacción del Triacetil-quitotriosa con la WGA mediante STD

Para el estudio de la interacción, se preparó una muestra del Triacetil-quitotriosa (1mM) en presencia de WGA (en relación molar 100:1) y se registraron los espectros de STD a dos frecuencias de saturación: 100 ppm (frecuencia *off*-resonance) y 7 ppm (frecuencia *on*-resonance en la zona aromática donde el ligando no presenta ningún protón no intercambiable) a pH 6 y un tiempo de saturación de 2s en un espectrómetro Bruker Avance 500 MHz a 25°C en D₂O.

Asignación del GMDP

De una solución Stock del GMDP en D₂O a 11 mM, se preparo una muestra del GMDP solo a 2mM en ausencia de proteína para asignar sus correspondientes señales de resonancia mediante una combinación de experimentos TOCSY (20 y 60 ms), NOESY (400 y 800 ms) y ROESY (a 400 ms) a 298K en D₂O a pH 6 en un espectrómetro Bruker Avance 600 MHz.

Estudio de interacción del GMDP con la WGA mediante STD

Para el estudio de la interacción, se preparó una muestra del GMDP (2 mM) en presencia de 0,2 mM WGA) y se registraron los espectros de STD a dos frecuencias de saturación (tiempo de saturación = 2s; frecuencia *off*-resonance: 100 ppm; frecuencia *on*-resonance: 7 ppm).

Asignación homonuclear del dominio LysM

Con una muestra del dominio LysM purificada a 0,8 mM, se realizaron los espectros TOCSY con secuencias estándar y con tiempos mezcla de 20, 40, 60 y 70 ms para el

TOCSY y de 100 y 200 ms para el NOESY a 298 K en un equipo Bruker Avance 500 MHz.

Titulaciones de LysM-CAM1 con el pentaacetil-quitopentaosa

De una solución concentrada del pentaacetil-quitopentaosa a 9 mM (conteniendo el dominio LysM-CAM1 a la misma concentración inicial de 0,4 mM para mantener su concentración constante durante toda la valoración), se añaden sucesivos alícuotas sobre una muestra de LysM-CAM-1 a concentración constante (0,4 mM, en tampón de fosfato sódico con 10% de agua deuterada a pH 5,8) y se registraron 10 espectros de protón monodimensionales con concentraciones crecientes del quitopentasacárido [0,0, 0,21; 0,40; 0,66; 0,99, 1,58; 2,37; 3,30, 4,43 y 9,0 mM] en un espectrómetro Bruker Avance 500MHz a 298 K.

Titulaciones de LysM-CAM1 con resina de quitina

A partir de resina comercial compuesta por partículas de quitina (resina de quitina) fácilmente manipulables en su forma hidratada se añadieron alícuotas sucesivos de una suspensión de la resina a una disolución de LysM-CAM-1 a 0.4 mM, en tampón fosfato sódico a pH 5,8 y se adquirieron espectros de ^1H RMN tras cada adición en un espectrómetro Bruker Avance de 500MHz a 298K.

Titulaciones de LysM-CAM1 con el GMDP

De una disolución de GMDP (11 mM), se realizaron una serie de adiciones sucesivas de alícuotas sobre una muestra de LysM-CAM-1 a 0.4 mM a 298 K y pH 5,8 hasta llegar a una relación 5:1 GMDP:LysM adquiriendo espectros ^1H RMN tras cada adición en un espectrómetro Bruker Avance de 500MHz.

Titulaciones de LysM-CAM1 con PGN

Se suspendieron 6,3 mg de PGN extraído de la pared de *E. coli* en 2 ml de D_2O . Una alícuota de esta suspensión se diluyo con D_2O hasta una concentración final de 0,8 mg de peptidoglicano por mililitro y se utilizó para realizar sucesivas adiciones de 20 μl cada una sobre la muestra de LysM-CAM-1 (0.4 mM, pH 5,8) hasta alcanzar una concentración final de 0,18 mg/ml de peptidoglicano en el tubo de RMN. Tras cada

adición, se adquirieron espectros de ^1H RMN en un espectrómetro Bruker Avance de 500MHz a 298K.

Asignación homonuclear del dominio PASTA

Con una muestra del dominio PASTA a 0.8 mM en PBS 10% D_2O a pH 7.3, se adquirieron los espectros TOCSY a 60 ms y NOESY a 200 ms en un espectrómetro Bruker Avance 600 MHz a 298K.

Asignación heteronuclear del dominio PASTA

A partir de una muestra del dominio PASTA a 0.6 mM marcado en ^{15}N se adquirieron los experimentos ^1H - ^{15}N -HSQC, TOCSY-HSQC y NOESY-HSQC en un espectrómetro Bruker 600 MHz a 298K en un buffer PBS a 10% D_2O y a pH 7.3.

Con otra muestra del mismo dominio pero esta vez con doble marcaje isotópico (^{13}C y ^{15}N) a 0.9 mM y en tampón fosfato sódico 10% D_2O a pH 6.5, se adquirieron los espectros, HNCA, HN(CO)CA, CBCA(CO)NH y HNCACB en un aparato de 800 MHz del Instituto Rocasolano de Química-Física del CSIC en Madrid, en colaboración con la Dra. Marta Bruix.

Asignación homonuclear del muropéptido

Se suspendieron 2 mg de muropéptido extraído de la pared de *B. subtilis* en $\frac{1}{2}$ ml de H_2O con 10% de D_2O , y se adquirieron los espectros TOCSY a 60 ms y NOESY a 350 ms en un espectrómetro Bruker Avance 600 MHz a 298K.

Titulaciones del dominio PASTA con el muropéptido

El muropéptido suspendido en $\frac{1}{2}$ ml de H_2O con 10% de D_2O y se utilizó para realizar sucesivas adiciones sobre la muestra de LysM-CAM-1 a 0,4 mM hasta alcanzar una relación de 20:1 Muropéptido:dominio-PASTA, adquiriendo los espectros de ^{15}N -HSQC tras cada adición en un espectrómetro Bruker 600 MHz a 298K.

Titulaciones del dominio PASTA con la ampicilina

De una solución concentrada de la ampicilina a 20 mM, se añadieron alícuotas sobre una muestra del dominio PASTA a 0,2 mM hasta un exceso 10:1 del antibiótico y se adquieren los espectros de ^{15}N -HSQC tras cada adición en un espectrómetro Bruker 600 MHz a 298K.

Determinación de la K_d de la interacción de la Gal con la VAA

En muestras que contenían una concentración fija de la VAA a 50 μM , se realizaron una serie de experimentos de titulación a diversos tiempos de saturación (0.5, 1, 2, 4, 8 s) y a diferentes concentraciones de ligando [1, 1.26, 2, 3.6, 6.1 y 9.8 mM], empleando la técnica de STD-NMR en un espectrómetro Bruker Avance de 500MHz a 298K.

Determinación de la K_d de la interacción de la α y β -Metil-Gal con la VAA

Análogamente a lo que se aplicó a la galactosa, se realizaron experimentos de STD-NMR con cada uno de los metil derivados independientemente a varias concentraciones de ligando y a distintos tiempos de irradiación de la viscumina (0.5, 1, 2, 4, 8 s) la que se mantiene en ambos casos durante todo el proceso de la titulación a una concentración fija equivalente a 50 μM en un espectrómetro Bruker Avance de 500MHz a 298K.

Competición entre el α - y β -Metil-Galactosido para la unión a la Viscumina

Se adquirieron una serie de experimentos de STD-RMN con concentraciones crecientes de β -metil galactosido [0, 0.3, 0.6, 1.2, 2.1, 3.9, 8.2 mM] sobre una concentración fija de α -metil galactosido a 1 mM en presencia de viscumina a 50 μM en un espectrómetro Bruker Avance de 500MHz a 298K.

2.2.2 Procesado de los espectros

Los espectros mono y bidimensionales se procesaron con el programa TopsSpin 2.0 (Bruker Biospin, Karlsruhe, Alemania), a los cuales se les aplicó la transformada de Fourier y se realizó la corrección de línea base en la dimensión directa e indirecta según fuera necesario en cada caso.

2.2.3 Aplicación de RMN en la determinación de la estructura 3D de polipéptidos y proteínas

Todos los espectros se procesaron con el software TOPSPIN y se analizaron y asignaron usando el programa CARA (Computer Aided Resonance Assignment, <http://cara.nmr.ch/>) diseñado y desarrollado por Rochus L.J. Keller en el grupo del Dr. Kurt Wüthrich en el ETH de Zürich, Suiza [5].

Asignación Homonuclear

Para la determinación de la estructura tridimensional de moléculas mediante RMN, en primer lugar, es necesario realizar la asignación inequívoca de los desplazamientos químicos (δ) de las señales de resonancia de cada uno de los átomos magnéticamente activos de la molécula en estudio. Esta asignación no es fácil en el caso de las proteínas grandes, ya que los δ de los protones de la molécula (que pueden ser centenares o incluso miles) se distribuyen en un rango de unas pocas ppm produciéndose un fuerte solapamiento de las señales. Es por tanto, necesario recurrir a experimentos bidimensionales (y multidimensionales) que permiten expandir el rango de observación de las señales de RMN a dos o más dimensiones amplificándose, por tanto, la dispersión de las señales. Adicionalmente estos experimentos bi-multidimensionales permiten establecer correlaciones entre distintos núcleos de la molécula y que a su vez dichas correlaciones se pueden transformar en información estructural.

En proteína pequeñas sin enriquecimiento isotópico o etiquetadas uniformemente con ^{15}N , se aplicó la estrategia de asignación desarrollada por Wüthrich y col. El primer paso es la identificación de los sistemas de espín correspondientes a los distintos tipos de aminoácidos a partir de las conectividades debidas a los acoplamientos escalares. Estas conectividades se detectan en los experimentos TOCSY, o 3D ^1H - ^{15}N TOCSY-HSQC

En el experimento TOCSY la magnetización fluye a lo largo de toda la red de acoplamientos, y se obtienen señales de correlación entre todas las frecuencias de los protones que pertenecen a un mismo sistema de espín [6],[7]. Los planos a cada δ de ^{15}N de un experimento 3D ^1H - ^{15}N TOCSY-HSQC contienen las correlaciones TOCSY de los sistemas de espín cuyo N-amídico tiene ese valor de δ y, evidentemente, el

solapamiento es mucho menor que en un 2D TOCSY. Así pues, si el solapamiento lo permite, se pueden agrupar las resonancias presentes en sistemas de espines de desplazamientos químicos conocidos, que constituyen patrones característicos. Los sistemas de espines de los 20 aminoácidos naturales se pueden agrupar en 8 grupos con patrones distintivos. Los residuos de Ala, Thr, Gly, Ser y Val constituyen grupos propios y sus sistemas de espines pueden identificarse directamente con el aminoácido al que pertenecen, mientras que los de Asp, Asn, Cys, Tyr, Phe, His, Trp constituyen el grupo AMX (las resonancias aromáticas no están acopladas escalarmente al sistema que parte del HN). Finalmente, Glu, Gln, Arg, Lys, Met, Leu e Ile forman parte de los sistemas de espín de “cadena larga”. Las Pro, al no presentar HN, constituyen un caso aparte, ya que la identificación de su sistema de espines es más complicada. Deben asignarse a partir del H α , que se encuentra en una región espectral más colapsada.

Normalmente, en proteínas se suelen registrar varios espectros TOCSY con diferentes valores de tiempo de mezcla. Usando entre 20-40 ms se obtienen los picos de cruce de los espines directamente acoplados, mientras que con tiempos mayores pueden llegar a espines situados relativamente alejados en la cadena.

Una vez identificadas todas las señales de resonancia, el segundo paso consiste en la identificación de cada sistema de espín con un residuo concreto de la secuencia de la proteína. Mediante experimentos tipo NOE: Nuclear-Overhauser-Effect [8], se podrá observar al menos una señal de correlación entre dos residuos consecutivos, que permitirá la conexión secuencial de pares de sistemas de espines [8]. Independientemente de la conformación del esqueleto polipeptídico siempre es posible encontrar una distancia secuencial (entre aminoácidos contiguos en la secuencia) inferior a 3 Å entre el HN de un residuo y los protones HN, H α o H β s del residuo anterior (Figura 2.2). Por tanto, en los experimentos 2D NOESY y 3D ¹H-¹⁵N NOESY-HSQC A partir del establecimiento de pares, tríos, y, en general, segmentos de sistemas de espines consecutivos en la secuencia y comparando la naturaleza de esos sistemas de espines con la secuencia de la proteína, se puede asignar cada sistema identificado a un aminoácido concreto de la secuencia.

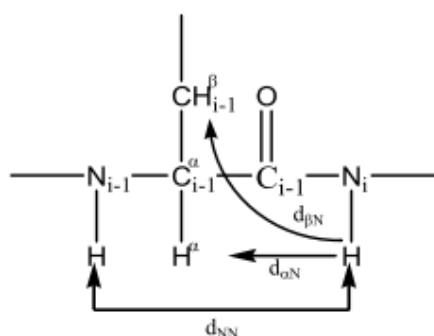


Figura 2.2 Definición de las conexiones secuenciales de un fragmento de cadena peptídica. Las flechas señalan las tres conexiones secuenciales interprotonicas y sus distancias características d_{NN} , $d_{\alpha N}$ y $d_{\beta N}$.

Asignación Heteronuclear

La asignación de proteínas de peso molecular mayor de 15 kDa sólo es posible, de forma general, si se dispone de ellas etiquetadas con ^{15}N y ^{13}C o con ^{15}N , ^{13}C y ^2H . En estas proteínas doble o triplemente etiquetadas, la conexión secuencial entre los sistemas de espín se basa en las conectividades observadas en espectros 3D y/o 4D en los que se hace uso de acoplamientos escalares a uno o dos enlaces en los que participan los heteronúcleos (Figura 2.3). Este método fue desarrollado por Bax y colaboradores [9] y presenta ventajas respecto a los métodos anteriores, basados en el efecto NOE. Ahora, ya que los acoplamientos escalares empleados son poco dependientes de la conformación, se evitan errores de asignación debidos a una interpretación incorrecta de los NOEs secuenciales, cuyas intensidades sí son dependientes de la conformación.

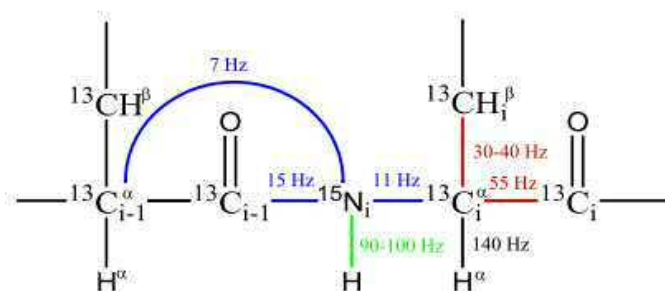


Figura 2.3 Constantes de acoplamiento a uno y dos enlaces involucradas en la transferencia de magnetización en los experimentos de triple resonancia. Las constantes de acoplamiento ^{15}N - ^{13}C se señalan en azul, las ^{13}C - ^{13}C en rojo, las ^{15}N - ^1H en verde y las ^{13}C - ^1H en negro.

Las estrategias de asignación a través de enlaces covalentes son múltiples, ya que se pueden combinar distintos experimentos 3D y 4D para obtener las conexiones secuenciales entre residuos contiguos en la secuencia (Tabla 2.1). El proceso comienza con la asignación de los núcleos del esqueleto peptídico, ^1HN , $^1\text{H}^{15}\text{N}$, ^{13}CO , $^1\text{H}\alpha$, $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ y $^1\text{H}\beta$, y posteriormente se completa con la asignación de las cadenas laterales de los distintos aminoácidos.

El uso combinado de los experimentos 3D HNCA y 3D HN(CO)CA es una de las posibilidades más utilizadas para realizar la asignación secuencial.

Tabla 2.1 Conectividades intrarresiduales y secuenciales observadas en algunos experimentos tri-y tetra-dimensionales, útiles para la asignación secuencial a través de enlaces covalentes.

Experimento	Conectividad intrarresidual	Conectividad interresidual
3D HNCA	$^1\text{H}_{\text{N}(i)}-^{15}\text{N}_{(i)}-^{13}\text{C}_{\alpha(i)}$	$^1\text{H}_{\text{N}(i)}-^{15}\text{N}_{(i)}-^{13}\text{C}_{\alpha(i-1)}$
3D HN(CO)CA		$^1\text{H}_{\text{N}(i)}-^{15}\text{N}_{(i)}-^{13}\text{C}_{\alpha(i-1)}$
3D H(CA)NH	$^1\text{H}_{\alpha(i)}-^{15}\text{N}_{(i)}-^1\text{H}_{\text{N}(i)}$	$^1\text{H}_{\alpha(i)}-^{15}\text{N}_{(i+1)}-^1\text{H}_{\text{N}(i+1)}$
3D HN(CA)CO	$^1\text{H}_{\text{N}(i)}-^{15}\text{N}_{(i)}-^{13}\text{CO}_{(i)}$	$^1\text{H}_{\text{N}(i)}-^{15}\text{N}_{(i)}-^{13}\text{CO}_{(i-1)}$
3D HNCO		$^1\text{H}_{\text{N}(i)}-^{15}\text{N}_{(i)}-^{13}\text{CO}_{(i-1)}$
3D HCACO	$^1\text{H}_{\alpha(i)}-^{13}\text{C}_{\alpha(i)}-^{13}\text{CO}_{(i)}$	
3D HCA(CO)N		$^1\text{H}_{\alpha(i)}-^{13}\text{C}_{\alpha(i)}-^{15}\text{N}_{(i+1)}$
4D CBCANH	$^{13}\text{C}_{\beta(i)}-^{13}\text{C}_{\alpha(i)}-^{15}\text{N}_{(i)}-^1\text{H}_{\text{N}(i)}$	$^{13}\text{C}_{\beta(i)}-^{13}\text{C}_{\alpha(i)}-^{15}\text{N}_{(i+1)}-^1\text{H}_{\text{N}(i+1)}$
4D CBCA(CO)NH		$^{13}\text{C}_{\beta(i)}-^{13}\text{C}_{\alpha(i)}-^{15}\text{N}_{(i+1)}-^1\text{H}_{\text{N}(i+1)}$

El experimento HNCA correlaciona los δ de ^1H y ^{15}N amídicos con el δ de $^{13}\text{C}\alpha$ de un aminoácido dado, haciendo uso de la constante de acoplamiento directa $^{15}\text{N}-^{13}\text{C}\alpha$ a un enlace (~ 11 Hz). Asimismo, este experimento proporciona correlaciones secuenciales por transferencia de magnetización desde el ^{15}N de un aminoácido al $^{13}\text{C}\alpha$ del residuo que le precede en la secuencia, a través de la constante de acoplamiento $^{15}\text{N}-^{13}\text{C}\alpha$ a dos enlaces (~ 7 Hz). Por tanto, en este experimento, para cada HN se obtienen dos picos de correlación en la dimensión de carbono, uno que corresponde a la correlación intrarresiduo $^1\text{H}_{\text{N}(i)}-^{15}\text{N}_{(i)}-^{13}\text{C}_{\alpha(i)}$ y otro, a la interresidual $^1\text{H}_{\text{N}(i)}-^{15}\text{N}_{(i)}-^{13}\text{C}_{\alpha(i-1)}$ (Figura 2.4).

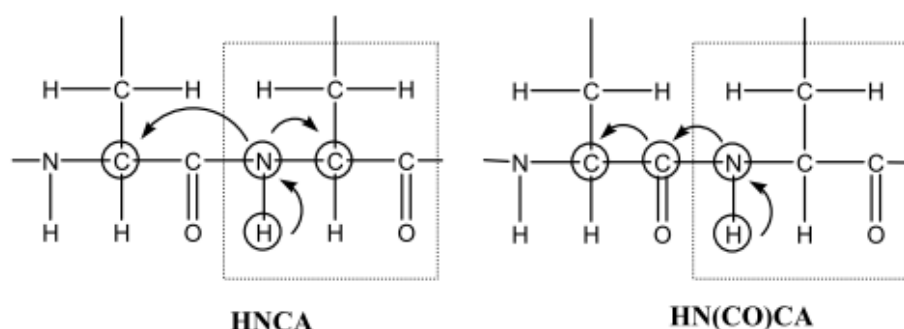


Figura 2.4 Núcleos implicados en la transferencia de magnetización en los experimentos de triple resonancia, HNCA y HN(CO)CA. Los átomos recuadrados pertenecen al mismo residuo.

El experimento HN(CO)CA permite obtener correlaciones secuenciales de los δ de ^1H y ^{15}N amídicos de un aminoácido con el δ de $^{13}\text{C}\alpha$ del aminoácido que le precede en la secuencia. La transferencia de magnetización se produce a través de las constantes de acoplamiento J_{NCO} y J_{CaCO} . Estas correlaciones secuenciales también aparecen en el HNCA pero en ocasiones no son fáciles de distinguir de las correlaciones intrarresiduo en este experimento puesto que las constantes de acoplamiento $1J_{\text{CaN}}$ y $2J_{\text{CaN}}$ pueden ser de la misma magnitud, o los δ de $^{13}\text{C}\alpha$ intrarresiduo e interresiduo pueden coincidir. Por tanto, en estos casos, es muy útil disponer del experimento HN(CO)CA, que sólo presenta las correlaciones interresiduales. Además, la sensibilidad del experimento HN(CO)CA es mayor que la del HNCA para la detección de correlaciones interresiduales, porque la transferencia de magnetización a través de las constantes de acoplamiento a un enlace J_{NCO} y J_{CaCO} es más eficiente que la transferencia vía la constante a dos enlaces $2J_{\text{CaN}}$ que es relativamente pequeña.

La asignación secuencial de los NH del esqueleto también puede realizarse mediante la combinación de los experimentos 3D HNCO y 3D HN(CA)CO. El experimento HNCO permite obtener la correlación secuencial de los δ de ^1H y ^{15}N amídicos de un aminoácido con el δ de ^{13}C del carbonilo del aminoácido que le precede en la secuencia. El experimento HN(CA)CO proporciona, además de la correlación secuencial, la correlación de los δ de ^1H y ^{15}N amídicos con el δ de ^{13}CO del mismo residuo. Los núcleos implicados en la transferencia de magnetización se muestran en la (Figura 2.5)

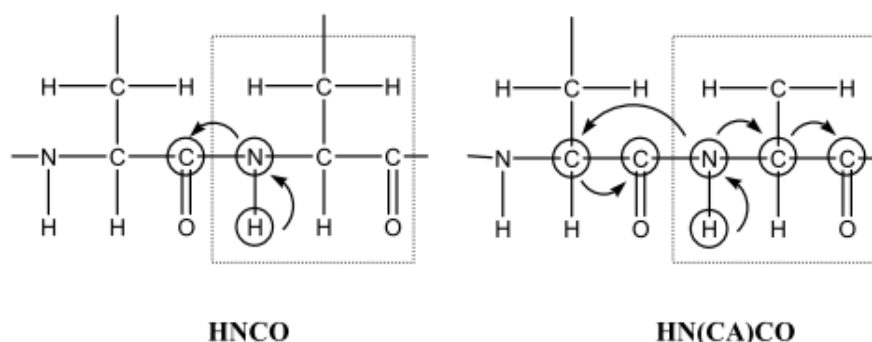


Figura 2.5 Núcleos que participan en la transferencia de magnetización en los experimentos de triple resonancia HNCO y HN(CA)CO. Los átomos recuadrados pertenecen al mismo residuo.

Calculo de estructura

El cálculo de la estructura tridimensional de una proteína consiste en un protocolo que permite encontrar un conjunto de estructuras consistentes con los datos de RMN y con el conocimiento químico a priori de las biomoléculas (geometría covalente y restricciones estéricas). El método de cálculo debe ser capaz de ajustar todas las restricciones experimentales a una geometría y de explorar adecuadamente el espacio conformacional. Para ello, se emplean métodos de geometría de distancias y de dinámica molecular restringida (normalmente, usando una estrategia de temperatura variable, también denominada “enfriamiento simulado”. Mediante geometría de distancias, los primeros procedimientos de cálculo hicieron uso de la matriz métrica, operando en el espacio de distancias. La matriz de distancias ideal contendría las distancias entre todos los átomos de la biomolécula y, a partir de ella, se obtendrían las coordenadas cartesianas. En la práctica, no se conocen todos los términos de la matriz, y se utilizan dos matrices, una con las cotas superiores de las distancias y otra con las inferiores. El inconveniente de estos métodos es la dificultad de incluir restricciones experimentales diferentes de las de distancia proporcionadas por los NOEs, como son las restricciones angulares proporcionadas por las constantes de acoplamiento o los desplazamientos químicos. Los métodos más empleados en la actualidad que se clasifican dentro del protocolo de geometría de distancias, aunque en rigor no lo sean, se basan en la minimización de una función blanco variable, que opera generalmente en el espacio de los ángulos de torsión. La función blanco es cero si se cumplen todas las restricciones experimentales y distinta de cero en cualquier otro caso. Dicha función se

define con las propiedades adecuadas para ser derivable y continua, de modo que el problema de obtener una estructura compatible con las restricciones experimentales se traslada al de minimización de una función. El algoritmo usado por el programa CYANA [10], que es el utilizado en esta Tesis, opera en el espacio de los ángulos de torsión, en torno a enlaces sencillos con giro no restringido, y mantiene fijos las distancias y ángulos de enlace, así como los ángulos de torsión en torno a enlaces con rotación impedida, de acuerdo a la geometría covalente estándar. De este modo, la dimensionalidad del problema se reduce en casi un orden de magnitud al ser los ángulos señalados los únicos grados de libertad considerados, y la geometría covalente se mantiene durante el cálculo. El problema consiste, por tanto, en encontrar los ángulos de torsión que hacen mínima la función de blanco variable.

Los métodos de mecánica molecular se fundamentan en el empleo de una función de energía potencial, que evalúa las características energéticas de la proteína en función de su conformación y, consecuentemente, permite refinar la estructura de acuerdo a dichos criterios energéticos. El refinamiento de una estructura CYANA dada puede lograrse por técnicas de minimización energética, o por métodos de dinámica molecular. Estos últimos se basan en la integración numérica de las ecuaciones del movimiento de un sistema de partículas, desarrolladas por Newton. Las posiciones iniciales provienen de una estructura de partida, resultado de un cálculo anterior, mientras que las velocidades se toman de una distribución de Maxwell-Boltzmann a una temperatura dada, normalmente aquella a la que se va a realizar la simulación. La forma general de un potencial V normalmente utilizado para cálculos de dinámica molecular es:

$$V = V_{\text{enlaces}} + V_{\text{ángulos}} + V_{\text{diedros}} + V_{\text{van der Waals}} + V_{\text{electrostática}}$$

En el protocolo de refinamiento de estructuras de RMN, se usan métodos de dinámica molecular restringida (DMR), en los que, al potencial que se acaba de describir, se le añaden unos términos de pseudo-energía, que penalizan las desviaciones de las restricciones experimentales de RMN. La ventaja principal de la dinámica molecular sobre la minimización de energía reside en su capacidad para saltar barreras de potencial y, por tanto, poder superar mínimos locales, con lo que el espacio conformacional accesible se muestrea más exhaustivamente. Para ello, el cálculo suele realizarse a altas temperaturas, que permiten un salto más eficiente de las barreras, para luego enfriar el sistema lentamente. Después de la simulación de dinámica molecular

suele llevarse a cabo una minimización de energía en el campo de fuerzas considerado. Esta minimización final proporciona estructuras con unas características energéticas mejoradas y elimina la energía cinética introducida en el sistema.

2.2.4 Aplicación de RMN en el reconocimiento molecular carbohidrato/proteína

2.2.4.1 Métodos basados en la observación del ligando

Diferencia de Transferencia de Saturación (STD)

Los experimentos de STD se llevaron a cabo según el método propuesto por Meyer, Peters y col. [11] (ver Introducción General). Los efectos STD se adquirieron en los espectrómetros de 500 y 600 MHz y se evaluaron mediante la substracción de dichos espectros.

- Calculo de K_d mediante STD

Estudios recientes [11-19], han propuesto la utilización de los experimentos de STD para la determinación directa de constantes termodinámicas de equilibrio en procesos de formación de complejos ligando-proteína mediante valoraciones directas o por estudios de competición entre ligandos.

Para poder calcular constantes de equilibrio a partir de datos de transferencia de saturación es necesario controlar de manera precisa las condiciones experimentales y así mismo definir la relación entre la cinética de la transferencia de saturación la relajación de los espines y la constante de equilibrio.

Para todos los protones de cada ligando, se construyeron las curvas de crecimiento de saturación a diferentes concentraciones, Estas curvas se ajustaron a una función mono-exponencial asintótica dada por la ecuación:

$$\text{STD}(t_{\text{sat}}) = \text{STD}_{\text{max}} \times (1 - e^{-(k_{\text{sat}} \times t_{\text{sat}})})$$

Donde t_{sat} es el tiempo de saturación, mientras que STD_{max} y k_{sat} son parámetros ajustables. STD_{max} corresponde al valor máximo de STD que un protón dado puede alcanzar y k_{sat} tiene forma de constante de velocidad de saturación que refleja a la velocidad con la que se alcanza el estado estacionario en la curva de saturación. Esta

constante se puede asimilar a una constante de velocidad macroscópica que incluye tanto la velocidad de transferencia de saturación desde los espines saturados de la proteína a los espines del ligando cuando esta unido a la proteína así como la velocidad de relajación longitudinal de dichos espines del ligando cuando no están en contacto con la proteína y la posible reasociación del ligando ya saturado que se va acumulando en el estado no asociado y que vuelve a ocupar el sitio de unión en el receptor. Después del ajuste matemático, la pendiente inicial del STD, que es una medida de la velocidad inicial de saturación (V_0) y donde la relajación de los espines del ligando en la forma libre y la posible reasociación están minimizados, es obtenida como producto de dos parámetros; $STD_{max} \times k_{sat}$

$$V_0 = STD_{max} \times k_{sat}$$

Las velocidades iniciales de saturación para cada protón, en semejanza a un proceso enzimático, son proporcionales a la concentración de forma unida [Proteína.Ligando] y permiten la estimación de las constantes de disociación mediante su ajuste a ecuaciones hiperbólicas del tipo de isothermas de Langmuir:

$$V_{i_{obs}} = \frac{(V_{i_{max}} * L)}{(L + K_d)}$$

donde, $V_{i_{obs}}$ es la velocidad inicial de saturación observada para cada protón a una concentración, L , del ligando en forma no asociada, $V_{i_{max}}$, la velocidad máxima de saturación a concentración constante de ligando y K_d la constante de disociación.

Por último, el crecimiento inicial de STD (V_0) sobre la concentración del ligando se ajustó para cada protón por la ecuación de Michaelis Menten para determinar la K_d y V_{max} . Para cada protón del ligando y a diferentes concentraciones, se representan las curvas de crecimiento del porcentaje de saturación referido a la concentración total del ligando presente en la muestra expresadas como concentración saturada de cada protón, obtenida por el producto del porcentaje de saturación multiplicado por la concentración de ligando total presente en la muestra ($\%STD \times [L_{total}]$).

- Experimentos de competición por STD

Un protocolo que permite estudiar la competencia entre los ligandos para el mismo receptor mediante el uso de experimentos STD. En esta tesis se aplicó por el estudio de las interacciones de los anómeros de la galactosa con la VAA. Utilizando un ligando como referencia llamado “espía” o “reportero” que debe ser una molécula con una constante de disociación conocida, es posible monitorizar la potencia inhibidora de un ligando con referencia al compuesto "espía", siguiendo la posible disminución de las intensidades de las señales STD del compuesto espía. Esta estrategia requiere que la K_d y la respuesta de la señal de RMN de la molécula espía sean previamente determinados. Esta estrategia ha sido aplicada previamente utilizando medidas de STD [20-21]. De hecho, el cribado en resonancia mediante experimentos de competencia por STD, utilizando una molécula espía bien caracterizada por el compuesto de ensayo, elimina las dos limitaciones importantes del método basado sobre la observación del ligando; los falsos negativos de alta afinidad de los compuestos ensayados (constantes de disociación en el rango nanomolar), y los falsos positivos resultantes de solubilidad limitada de los compuestos de ensayo y unión no específica al receptor.

La interferencia que causa cada compuesto en la unión de la molécula espía al receptor se sigue y se cuantifica por la disminución de la señal STD de la molécula espía. Y por tanto su potencia inhibidora se calcula a partir de la ecuación que define una constante de disociación aparente:

$$K'_{D,L} = K_{D,L} \left(1 + \frac{[I]_t}{K_{I,I}} \right)$$

Donde, $K_{i,i}$ es la constante de inhibición del inhibidor añadido, $K_{D,L}$, se refiere a la constante de disociación conocida (molécula espía), $[I]$ y $[L]$ son las concentraciones del inhibidor añadido y de la molécula espía respectivamente.

Las señales observadas de STD para el ligando **L** o molécula espía observada ($STD_{L,I}$) en presencia del otro compuesto o inhibidor **I** se relacionan con los valores de las constantes de disociación según la siguiente ecuación:

$$\frac{STD_{L,I}}{STD_L} = \frac{\left(\frac{STD_{L,max} * [L]_t}{K_{D,L} * (1 + \frac{[I]_t}{K_{D,I}}) + [L]_t} \right)}{\left(\frac{STD_{L,max} * [L]_t}{K_{D,L} + [L]_t} \right)}$$

- Modelización teórica de experimentos de STD por CORCEMA

Debido a que el método STD en RMN se basa sobre la transferencia de relajación transversal, métodos teóricos se han desarrollado para analizar esta última. Se ha descrito el programa CORCEMA basado sobre estrategias que permiten obtener los efectos cuantitativos del STD para un determinado modelo, considerando los procesos de relajación y las características cinéticas del sistema estudiado [15]. Este procedimiento considera la influencia del complejo y de la conformación del ligando sobre las velocidades de relajación de los protones del ligando, y luego, a sus correspondientes efectos de STD. El método CORCEMA también es útil en la determinación de las condiciones experimentales óptimas para la realización de las mediciones de STD-RMN en un complejo determinado mediante simulaciones por ordenador. También permite el refinamiento estructural de ligandos unidos a receptores determinados por modelado y cristalografía de rayos-x [14, 16]. Para la construcción de una matriz de relajación precisa, el protocolo CORCEMA requiere una estructura de alta resolución de la proteína.

El protocolo CORCEMA, fue empleado en el para la validación de los modelos 3D de complejos obtenidos a partir de los procedimientos de acoplamiento por AutoDock (figura 2.6). Este método permite la predicción de los efectos del STD para un conjunto de 100 complejos carbohidrato-proteína, una vez conocidos la constante de unión y los tiempos de correlación en estado libre y enlazado. Otros parámetros, tal como el tiempo de saturación, la concentración de ligando y de proteína, los protones metílicos en el ligando y en la proteína, así como el campo magnético, se registran como datos de entrada para el cálculo en el programa. Las intensidades teóricas de STD se comparan cuantitativamente a los obtenidos experimentalmente por RMN, por la función de calificación denominado NOE R-factor, tal como se aplica en el programa. Luego, se

conservan o rechazan las estructuras en función del NOE R-factor, para obtener el mejor ajuste entre los valores experimentales de STD-RMN y los valores predichos por el programa.

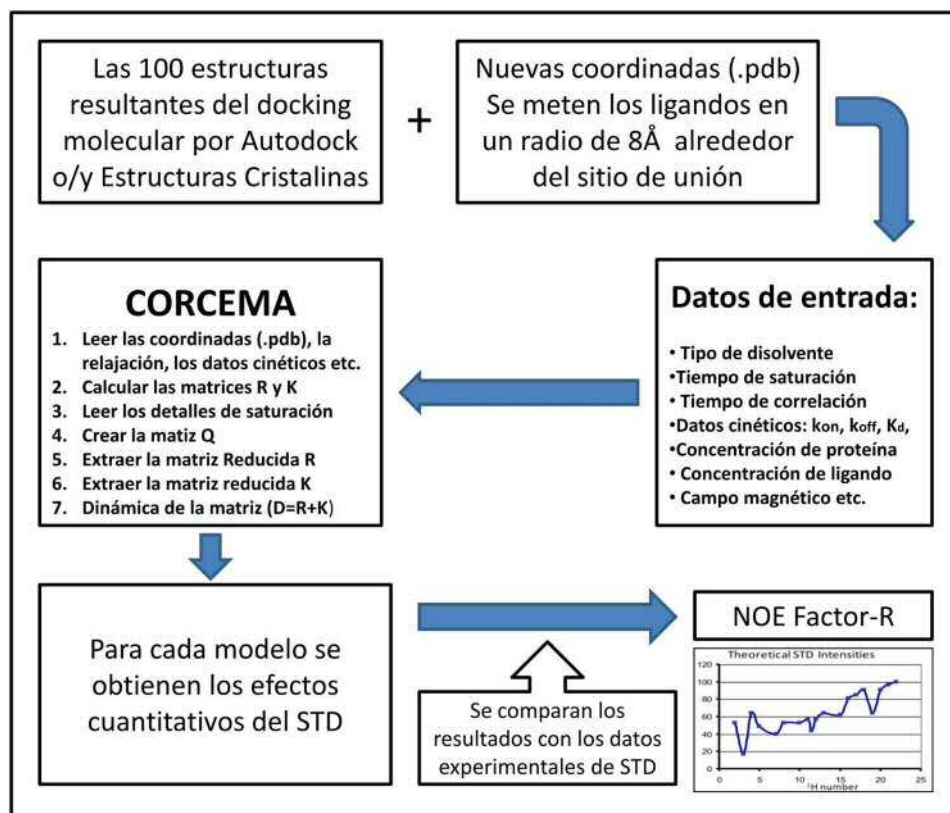


Figura 2.6 Esquema del protocolo para el cálculo teórico de las intensidades de STD y su comparación con los datos experimentales mediante el programa CORCEMA.

2.2.4.2 Métodos basados en la observación del receptor

Los parámetros espectroscópicos del receptor (proteína) se afectan al producirse la asociación con el ligando (carbohidrato), como por ejemplo, los desplazamientos químicos, constantes escalares y de difusión, velocidades de relajación, NOE, etc.

En esta tesis, se ha seguido la estrategia de observación de las perturbaciones de desplazamiento químico de las señales del receptor mediante experimentos de titulación de los ligandos sobre una concentración fija del receptor con objeto de cuantificar la constante de afinidad.

Método por perturbaciones de desplazamientos químicos del receptor

Las titulaciones con ligandos se llevaron a cabo utilizando dos experimentos de RMN; espectros monodimensional y homonucleares simples 1D ^1H -RMN y otros espectros 2D de correlación cuántica simple heteronucleares (HSQC) ^{15}N - ^1H -HSQC.

- Titulaciones mediante espectros ^1H RMN monodimensionales

Se calcularon los desplazamientos químicos en ppm de los residuos sobre el agregado de cantidades de crecientes de los ligandos utilizados

- Titulaciones mediante espectros ^{15}N -HSQC

Las titulaciones con ligandos se llevaron a cabo en experimentos separados, utilizando un espectro de correlación cuántica simple heteronuclear (HSQC).

Se utilizaron diferentes razones molares de proteína-ligandos, para seguir las variaciones de desplazamiento químico de las resonancias de los residuos perturbados por la interacción con el ligando.

2.3 TÉCNICAS DE MODELIZACION MOLECULAR

2.3.1 Estudio de acoplamiento o "docking" con AutoDock

Los cálculos de modelización molecular se realizaron con el programa AutoDock [22-23]. La preparación de los archivos de entrada y el análisis de los resultados se llevó a cabo con la interfase gráfica del programa, AutoDock Tools (ADT, operativa en Windows 2000, 98 y Linux).

Inicialmente, es necesario disponer de las coordenadas de la macromolécula (VAA) y del ligando. En este trabajo, se emplearon las coordenadas de la estructura obtenida por cristalografía (código pdb 1PUM). Los ligandos se dibujaron en el programa Chemdraw y se guardaron bajo la extensión .pdb

El primer paso en la preparación de los archivos de entrada de AutoDock 3.0 consiste en la adición de los hidrógenos (solo los polares para la proteína), las cargas y los parámetros de solvatación, al ligando y a la macromolécula. Las cargas de la proteína se adicionaron con ADT, el programa detecta sin problemas que la macromolécula es una

proteína y añade las cargas de Kollman (a partir de una librería de cargas de aminoácidos que tiene ADT).

Los archivos de coordenadas del ligando y de la macromolécula, así preparados, deben guardarse con las extensiones pdbqt. El archivo pdbqt, también lleva información acerca de las torsiones del ligando que van a poder girar libremente durante el cálculo, las torsiones de la proteína se mantienen fijas.

Los archivos (.pdbqt) se van a emplear en primer lugar para calcular un mapa de energía potencial y electrostática para cada tipo de átomo del ligando (*grid map*). Los mapas describen una caja tridimensional, centrada en algún punto de interés del receptor, que está formada por una red de puntos separados una distancia constante (espaciado de la red, *grid spacing*) (Figura 2.7). La magnitud del espaciado de la red varía entre 0.2 y 1 Å. Cada punto de la caja contiene la información de la energía resultante de la interacción de un átomo del ligando, situado en esa posición, con todos los átomos del receptor. En el ajuste, la posición de la proteína se mantiene fija, mientras que el ligando se va moviendo.

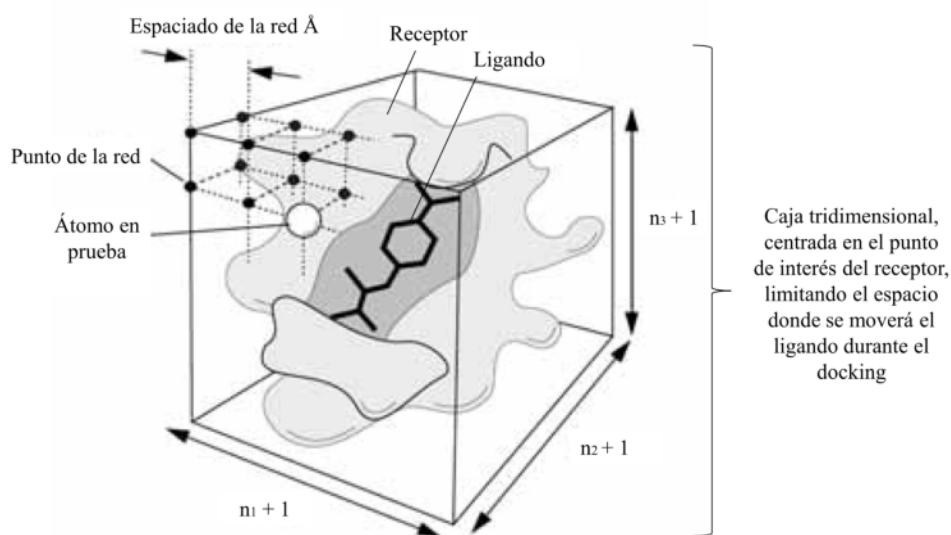


Figura 2.7 Imagen representativa de la caja 3D donde se realiza el proceso del Docking usando el programa Autodock. Imagen extraída y modificada del manual de Autodock [22-23].

Los mapas de la energía se calculan con el programa AutoGrid (operativo en SGI o Linux) incorporado en el paquete informático AutoDock 3.0. AutoGrid requiere un archivo (con extensión *gpf* generado con ADT) que especifica los parámetros del cálculo: tamaño de la caja, localización del centro, espaciado de la red y constante dieléctrica. En este trabajo conocemos los sitios de interacción de la proteína ya que esta previamente descritos en la estructura cristalina depositada en proteína *data bank*.

Los mapas de energía calculados junto con los archivos *.pdbqt* constituyen los archivos de entrada de AutoDock 3.0 (operativo en SGI o Linux). Al igual que con AutoGrid es necesario generar un archivo (con ADT) que lleve la información de las condiciones del cálculo, en este caso tiene extensión (*.dpf*). Método de ajuste que vamos a utilizar está basado en los Algoritmos Genéticos Lamarkianos (AGL). El método AGL permite analizar variaciones de la posición, orientación y conformación del ligando en la búsqueda de posibles sitios de unión a la proteína. Los parámetros utilizados se denominan "variables de estado" y, por analogía con la terminología genética, se han denominado genotipo del ligando. Las coordenadas que resultan del ajuste (docking) del ligando en el sitio de unión de la proteína y sus energías intramoleculares correspondientes serían el análogo del fenotipo. Cada ciclo de ajuste tiene una etapa de cruzamiento, una de mutación, una minimización energética y una etapa de selección. Inicialmente, el programa genera, de manera aleatoria, una población de estructuras del ligando en la que el número de individuos es elegido por el usuario. A continuación se produce el cruzamiento al azar de parejas de individuos, intercambiando dos de sus variables de estado. Las nuevas estructuras generadas reemplazan a las originales, de modo que el número total de individuos se mantiene constante. Posteriormente, se generan, de nuevo de modo aleatorio, mutaciones en las variables de estado y se realiza una minimización energética local. Las estructuras optimizadas reemplazan a las originales y, de este modo, las mejoras conseguidas con la minimización (que en este caso reflejan la adaptación del fenotipo al entorno, en términos evolutivos) se transmiten a las nuevas generaciones de individuos. Por ello, se denominan algoritmos lamarkianos. Las mejores estructuras así obtenidas se seleccionan para pasar al siguiente ciclo de ajuste. Mediante la concatenación de varios de estos ciclos (cruzamiento-mutación-minimización-selección) se genera una estructura que constituye un posible modo de unión en las condiciones del cálculo. El número de ciclos empleados para generar cada estructura se puede variar, cambiando el número de evaluaciones

energéticas. El proceso global se repite varias veces para generar distintos modos de unión. Las soluciones se agrupan de acuerdo con la desviación cuadrática media (RMSD) de unas con respecto a las otras. En los cálculos realizados en este trabajo, el tamaño de la población de estructuras generada es de 100 individuos y el número máximo de evaluaciones energéticas para los ciclos de cruzamiento-mutación-minimización-selección es de 3×10^6 . El proceso global se repite 100 veces, generando 100 posibles modos de unión. Las coordenadas de las estructuras del ligando generadas vienen dadas en un archivo (.dlg) que se puede visualizar con ADT.

3 REFERENCIAS BIBLIOGRÁFICAS

1. M. Dagert y S.D. Ehrlich, *Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells*. Gene, 1979. **6**(1): p. 23-28.
2. R. Anand, *PULSED FIELD GEL-ELECTROPHORESIS - A TECHNIQUE FOR FRACTIONATING LARGE DNA-MOLECULES*. Trends in Genetics, 1986. **2**(11): p. 278-283.
3. M.M. Bradford, *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding*. Anal Biochem, 1976. **72**: p. 248-254.
4. A. Bax y D.G. Davis, *Practical aspects of two-dimensional transverse NOE spectroscopy*. Journal of Magnetic Resonance (1969), 1985. **63**(1): p. 207-213.
5. R. Keller, *Optimizing the Process of Nuclear Magnetic Resonance Spectrum Analysis and Computer Aided Resonance Assignment*. 2004/2005, Swiss Federal Institute of Technology Zurich (ETH Zürich).
6. L. Braunschweiler y R.R. Ernst, *Coherence Transfer By Isotropic Mixing - Application To Proton Correlation Spectroscopy*. Journal of Magnetic Resonance, 1983. **53**(3): p. 521-528.
7. A. Bax y D.G. Davis, *Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy*. Journal of Magnetic Resonance, 1985. **65**(2): p. 355-360.
8. D. Neuhaus y M.P. Williamson, *The nuclear Overhauser effect in structural and conformational analysis*. 1989, New York: VCH Publishers.
9. M. Ikura, L.E. Kay y A. Bax, *A novel approach for sequential assignment of ¹H, ¹³C, and ¹⁵N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin*. Biochemistry, 1990. **29**(19): p. 4659-4667.
10. P. Guntert, *Automated NMR structure calculation with CYANA*. Methods Mol Biol, 2004. **278**: p. 353-378.
11. M. Mayer y B. Meyer, *Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor*. J Am Chem Soc, 2001. **123**(25): p. 6108-6117.
12. J. Yan, A.D. Kline, H. Mo, M.J. Shapiro y E.R. Zartler, *The effect of relaxation on the epitope mapping by saturation transfer difference NMR*. J Magn Reson, 2003. **163**(2): p. 270-276.
13. S. Kemper, M.K. Patel, J.C. Errey, B.G. Davis, J.A. Jones y T.D. Claridge, *Group epitope mapping considering relaxation of the ligand (GEM-CRL): including longitudinal relaxation rates in the analysis of saturation transfer difference (STD) experiments*. J Magn Reson, 2010. **203**(1): p. 1-10.
14. V. Jayalakshmi y N. Rama Krishna, *CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities*. J Magn Reson, 2004. **168**(1): p. 36-45.
15. V. Jayalakshmi y N.R. Krishna, *Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes*. J Magn Reson, 2002. **155**(1): p. 106-118.
16. V. Jayalakshmi, T. Biet, T. Peters y N.R. Krishna, *Refinement of the conformation of UDP-galactose bound to galactosyltransferase using the STD NMR intensity-restrained CORCEMA optimization*. J Am Chem Soc, 2004. **126**(28): p. 8610-8611.
17. C.A. Lepre, J.M. Moore y J.W. Peng, *Theory and applications of NMR-based screening in pharmaceutical research*. Chem Rev, 2004. **104**(8): p. 3641-3676.
18. J. Angulo, P.M. Enriquez-Navas y P.M. Nieto, *Ligand-receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates*. Chemistry, 2010. **16**(26): p. 7803-7812.

19. D.W. Neuhaus , M. P. , *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, . Wiley-VCH, Weinheim,, 2000.
20. K. Feher, P. Groves, G. Batta, J. Jimenez-Barbero, C. Muhle-Goll y K.E. Kover, *Competition saturation transfer difference experiments improved with isotope editing and filtering schemes in NMR-based screening*. J Am Chem Soc, 2008. **130**(50): p. 17148-17153.
21. R. Meinecke y B. Meyer, *Determination of the binding specificity of an integral membrane protein by saturation transfer difference NMR: RGD peptide ligands binding to integrin α IIb β 3*. J Med Chem, 2001. **44**(19): p. 3059-3065.
22. D.S. Goodsell y A.J. Olson, *Automated docking of substrates to proteins by simulated annealing*. Proteins, 1990. **8**(3): p. 195-202.
23. G.M. Morris, D.S. Goodsell, R. Huey y A.J. Olson, *Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4*. J Comput Aided Mol Des, 1996. **10**(4): p. 293-304.

CAPITULO 3

**MODELOS DE INTERACCIONES DE
PROTEÍNAS CON CARBOHIDRATOS
DERIVADOS DE
N-ACETILGLUCOSAMINA**

Capítulo 3: Modelos de interacciones de proteínas con carbohidratos derivados de N-acetilglucosamina

Los últimos progresos en glicobiología siguen revelando el papel esencial que juegan los oligosacáridos de la superficie celular en diversos procesos de reconocimiento biológico, incluyendo el reconocimiento intercelular, la adhesión, simbiosis, el crecimiento celular y la diferenciación [1-3]. El factor clave en estas funciones biológicas es el poder de reconocimiento molecular de una proteína hacia un carbohidrato. Las lectinas, como se ha descrito en el capítulo de introducción, son capaces de reconocer sacáridos específicos en la célula. Como resultado aparecen en

una gran variedad de bioactividades. Desde el punto de vista del carbohidrato, existen multitudes de formas de presentación de las cadenas sacarídicas, desde glicoconjugados sencillos con una o unas pocas unidades sacarídicas unidas a proteínas y lípidos, glicósidos o simples oligosacáridos, hasta estructuras poliméricas complejas que forman parte de las paredes celulares. En este último caso es interesante mencionar las paredes bacterianas, basadas en una estructura de peptidoglicano que consiste en múltiples cadenas oligosacarídicas compuestas por varias unidades del disacárido de repetición GlcNAc-MurNAc (Figura 3.1 b), que se entrecruzan mediante fragmentos peptídicos elongados desde los restos de ácido murámico. Otros ejemplos son las paredes de hongos y los exoesqueletos de insectos y crustáceos que tienen una estructura basada en la quitina, un polisacárido natural formado por moléculas de GlcNAc unidas entre sí mediante enlaces glicosídicos β -1,4 (Figura 3.1 a).

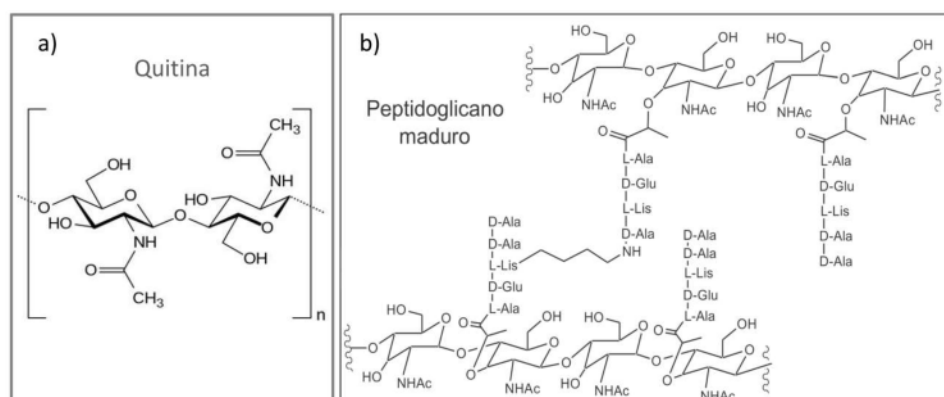


Figura 3.1 a) Representación de la unidad constituyente de la cadena de quitina. **b)** Estructura de un fragmento del peptidoglicano de la pared bacteriana tipo L-Lis.

Los organismos que interactúan con bacterias y hongos han desarrollado un amplio panel de proteínas dirigidas al reconocimiento de estructuras relacionadas con las paredes de estos microbios. Sus funcionalidades son muy diversas y incluyen la defensa frente a patógenos (ejemplo de la WGA o la heveína en plantas, que reconocen quitina y quitooligosacáridos) [4-6]. También permiten reconocer microorganismos parecidos como ocurre entre bacterias que usan proteínas que contienen los dominios LysM [7] o dominios PASTA [8] que reconocen el propio peptidoglicano y el de las bacterias vecinas como modo de control del estado de crecimiento y densidad de la población bacteriana. También median el reconocimiento de otros microorganismos beneficiosos o

simbiontes, donde volvemos a encontrar otra vez como ejemplo, a los mencionados dominios LysM, que también desempeñan el papel de receptor de factores de nodulación en plantas que reconocen lipoquitoligosacáridos liberados por bacterias simbiontes [3, 9].

1 DOMINIO HEVEÍNA

1.1 INTRODUCCIÓN Y ANTECEDENTES

Las plantas sintetizan una amplia gama de proteínas capaces de unirse reversiblemente a matrices compuestas por quitina y otros oligosacáridos estructuralmente relacionados a los lipoquitoligosacáridos. Los dominios heveína contienen un motivo estructural común de 30-43 aminoácidos con varias cisteínas y glicinas en posiciones altamente conservadas (Figura 3.3). Este motivo se reconoce como un dominio de unión a quitina con capacidad de mantener uniones reversibles. Aunque el término “proteínas de unión a quitina” indica la familia de proteínas que contienen uno o más dominios de unión a quitina. No obstante, la especificidad de unión de estas proteínas no se limita a la quitina, sino que puede extenderse a otros glicoconjugados complejos que contienen GlcNAc, NeuNAc o el peptidoglicano.

Los dominios de heveína están presentes en varias lectinas de manera aislada, como en la propia heveína y en su variante natural, la pseudoheveína. También se encuentran aislados en péptidos antimicrobianos (Ejemplo del péptido antimicrobiano Ac-AMP de *Amaranthus caudatus*), pero también se pueden encontrar en forma de repeticiones múltiples como en la WGA, con 4 repeticiones de un dominio heveína (Figura 3.2), o la *urtica dioica agglutinin* (UDA) con dos repeticiones. Este motivo de unión a quitina también se puede encontrar formando parte de proteínas con otras funcionalidades como las enzimas quitinasas de la clase I [10].

El dominio de heveína se ha convertido en una poderosa herramienta en varios campos de la investigación biológica y sus propiedades de lectina se han estudiado extensivamente en el campo de glicobiología [11]. Su actividad como proteína de defensa de plantas ha sido el enfoque de interés en biología vegetal [9][12]. Sus posibles aplicaciones en biotecnología se han enfocado a incrementar la resistencia hacia los

hongos, transformando las plantas con genes que codifican el dominio heveína [13-14] y en biomedicina como un agente antimicrobiano [15].

La proteína de unión a quitina con múltiples dominios heveína mejor caracterizada es la WGA presente en las semillas del trigo que reconoce oligosacáridos derivados de GlcNAc y residuos de NeuNAc en el extremo no reductor. Se han descrito sus propiedades antimicrobianas, ya que es capaz de inhibir el crecimiento de hongos a través de la interacción con los componentes de sus paredes celulares [5, 16-18]. Por otro lado, también es capaz de aglutinar células transformadas *in vitro* [19]. La WGA forma un homodímero de 36 kDa estable con un doble eje de simetría [20-22]. Cada cadena polipeptídica se presenta por cuatro dominios de heveína de 43 (A-D) en posiciones altamente conservadas y organizadas alrededor de un núcleo de cuatro enlaces disulfuro (Figura 3.2). Mediante cristalografía de rayos X y RMN se han caracterizado los sitios de unión a carbohidratos de complejos con la sialillactosa (Neu5Ac- α (2,3)-Gal- β (1,4)-Glc) [23] y sialoglicopéptidos [24] y otros disacáridos relacionados con la N-Acetilglucosamina, N,N'-diacetil quitobiosa (GlcNAc- β (1,4)-GlcNAc) [21, 25], GlcNAc- β (1,6)-Gal y GlcNAc- β (1,6)-Gal- β (1,4)-Glc [26]. Recientemente, mediante el uso de ligandos divalentes ha sido posible demostrar la funcionalidad simultanea de los ocho sitios de unión del dímero de WGA (cuatro sitios en cada monómero debido al eje de simetría doble) [27]. Desde el punto de vista del carbohidrato, en nuestro grupo de trabajo, se ha demostrado que las cadenas N-glicánicas de glicoproteínas pueden ser reconocidas por los dominios heveína, no solo por su extremo no reductor, sino también por las N-acetilglucosaminas del “core” de las cadenas de N-glicano [28].

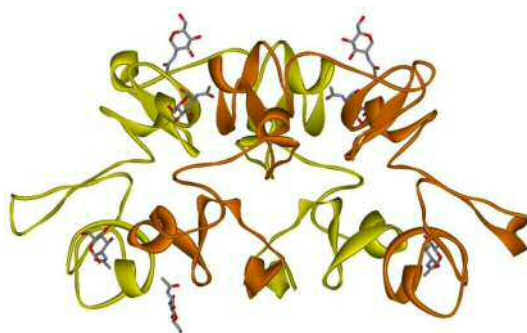


Figura 3.2 Imagen estructural del heterodímero de la proteína WGA en complejo con GlcNAc (código PDB; 2UVO).

Los anteriores estudios cristalográficos juntos con estudios estructurales por RMN de diversos dominios heveína en complejos con oligosacáridos de quitina han permitido determinar que, en estos dominios, los residuos aromáticos en las posiciones relativas (numeración según la secuencia de la heveína) 21, 23 y 30 se conservan en todos los dominios heveína (Figura 3.3) y juegan un papel clave en la unión con los carbohidratos. De hecho, permiten la estabilización del complejo a través de apilamientos carbohidrato-aromático, interacciones de van der Waals y enlaces de hidrógeno, como se muestra en la figura 3.4 [11]. El grupo hidroxilo de la serina 19 forma un enlace de hidrogeno con el carbonilo del grupo acetamido de una unidad de GlcNAc, y al mismo tiempo, el grupo metilo de este acetamido interactúa con el anillo aromático de la cadena lateral del residuo 30. También existe un enlace de hidrogeno entre el grupo hidroxilo de la Tyr 30 de la heveína con el OH-3 de la misma unidad de azúcar mencionada anteriormente.

		10	20	30	40	
Heveína	18:--EQCGRQAGGK	LCPNNLCCSQ	WGWCGSTDEY	CSPDHNCQSN	CKD 60	
Pseudoheveína	1:--EQCGRQAGGK	LCPNNLCCSQ	YGWCGSSDDY	CSPSKNCQSN	CKG 43	
WGA-B Isolectina I	69:--KRCGSQAGGA	TCTNNQCCSQ	YG YCGFGAEY	CGAGCQGGPC	RA- 111	
UDA-A	24:--QRCGSQGGGG	TCPQLWCCSI	WGWCGDSEPY	CGRTCENKCW	S-- 64	

Figura 3.3 Alineación de las secuencias de amino ácidos de la Heveína, Pseudoheveína, WGA dominio B y UDA dominio A. En negrita se indican los a.a que participan en el sitio de unión.

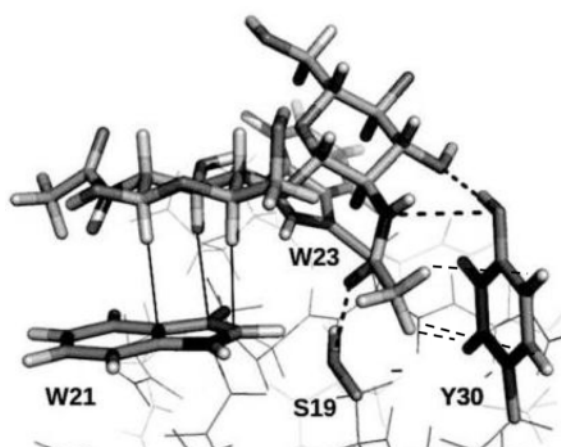


Figura 3.4 Interacciones proteína-carbohidrato que regulan el reconocimiento de la quitobiosa por la Heveína, tal como se determina por RMN. Las líneas discontinuas representan enlaces de hidrógeno que se establecen entre el azúcar y los residuos polares de la proteína (S19 y Y30). Las líneas continuas indican las interacciones CH- π .

Estudios previos [11, 18, 29-30] propusieron un sitio de unión extendido en el dominio heveína, capaz de alojar varias unidades de GlcNAc, permitiendo dos modos de unión para el triacetil quitotriosa (GlcNAc)₃.

Hay un primer modo de unión en el que se coloca el residuo GlcNAc terminal no reductor en el llamado subsitio +1, interaccionando con el Tyr 23 (CH- π), con la Ser 19 (enlace de hidrógeno) y con la Tyr 30 (CH₃-CH- π y enlace de hidrógeno). El residuo intermedio se encuentra en el subsitio +2, interaccionando con el Tyr 21 (CH- π), mientras que el residuo GlcNAc reducido proporciona muy pocos contactos con la lectina, en el subsitio +3 (Figura 3.5-a) [11].

En la segunda orientación, el residuo del extremo reductor se coloca en el subsitio +2, interaccionando con el Tyr 21, el residuo intermedio se coloca en el subsitio +1, interaccionando con el Tyr 23 (CH- π), Ser. 19 (enlace de hidrógeno) y con la Tyr 30 (CH- π y enlaces de hidrógeno), mientras que el extremo terminal no reductor muestra algunos contactos con la lectina en el llamado subsitio -1 (Figura 3.5-b) [5, 29].

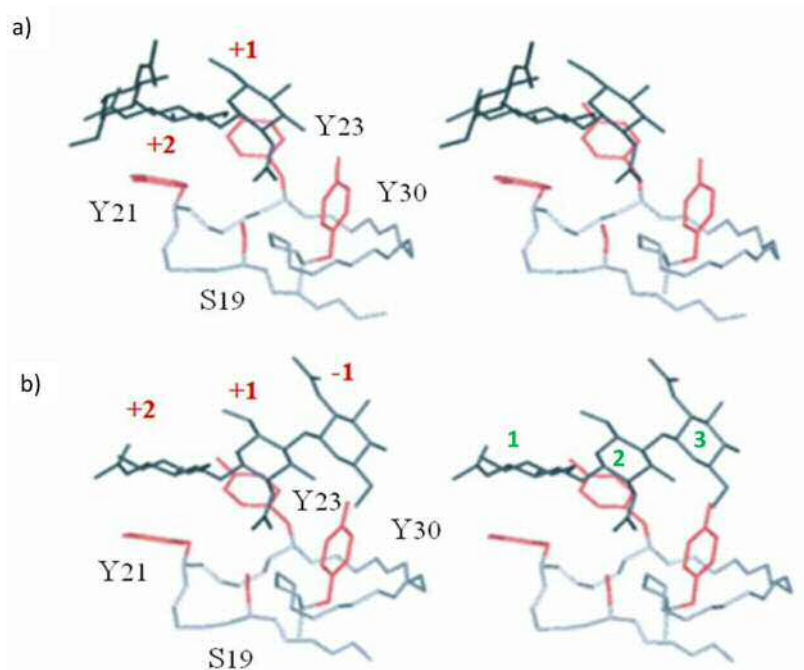


Figura 3.5 Representación estereoscópica de los dos modos de unión de la quitotriosa en el dominio B de la WGA descritos por Espinosa y col. [30] y determinados por RMN. El ligando se presenta en color verde numerándose los residuos desde el extremo reductor mientras que los anillos aromáticos de la proteína que participan en la interacción se presentan en color rojo numerándose tres subsitios de unión como -1, +1, +2.

Debido al hecho de que la quitina es el componente principal en la pared celular de los hongos y los exoesqueletos de los invertebrados, la función de defensa de las plantas y de las actividades antifúngicas de este tipo de péptidos se ha asociado con su carácter lectina como ligando de quitina [31]. Sin embargo, la descripción de las actividades antimicrobianas frente a otros microorganismos que no contienen quitina obliga a encontrar y caracterizar otros mecanismos de acción [31-32]. De hecho, en algunos casos, la actividad antimicrobiana se ha asociado al carácter policatiónico de algunos de estos péptidos [33]. Curiosamente, aunque el componente polisacárido del peptidoglicano bacteriano está muy relacionado con la quitina, hay escasos estudios que muestran la interacción directa entre oligosacáridos derivados de peptidoglicanos y proteínas con dominio heveína [6, 34] y no se han descrito datos de tipo estructural. Algunos ejemplos poco concluyentes han postulado una posible interacción. Por ejemplo, Lotan y col. [6] dedujeron que oligosacáridos complejos (di, tetra), derivados del peptidoglicano eran capaces de inhibir la hemaglutinación. Otros autores, como Ayoub y col. [4] mostraron otras evidencias de interacción en el caso de la aglutinina con dos dominios heveína extraída de la ortiga ("*Urtica Dioica Agglutinin*", UDA)

A pesar de esta escasez de información, hay que decir que la WGA se comercializa como un reactivo para histoquímica con aplicación para teñir bacterias gram-positivas (por ejemplo reactivo Alexa Fluor® 488 conjugate of WGA en MolecularProbes [35]), ya que muestra capacidad de reconocimiento del peptidoglicano bacteriano.

1.2 RESULTADOS Y DISCUSIÓN

El trabajo llevado en este apartado se ha centrado en el estudio estructural de la interacción de dominios de heveína con posibles ligandos relacionados con el peptidoglicano de las paredes bacterianas. El objetivo específico ha sido caracterizar, mediante técnicas de STD RMN desde el punto de vista del ligando, la interacción de fragmentos de peptidoglicano con dominios de heveína, e identificar la posible participación de la parte peptídica del peptidoglicano en la interacción.

Para tal fin, se eligió el GMDP (**G**lucosaminil-**M**uramil-**D**iPéptido, N-Acetil Glucosamina-N-Acetil murámico-L-Alanina-D-isoGlutamina, mostrado en la figura 3.6) como modelo de la unidad mínima disacáridica repetitiva de la estructura del peptidoglicano y la WGA como modelo de proteína con dominio heveína.

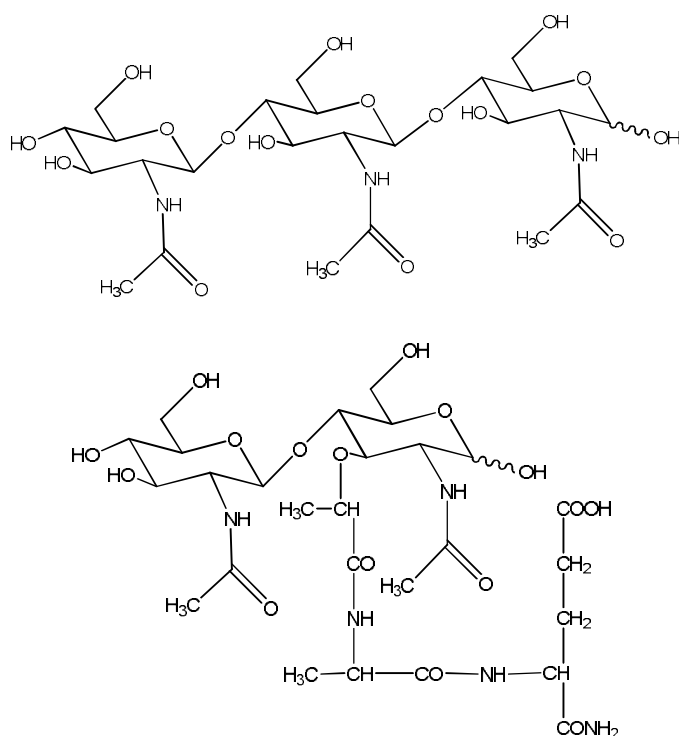


Figura 3.6 Estructura de la quitotriosa y del GMDP (N-Acetilglucosamina N-Acilmurámico L-Alanina D-Isoglutamina) ambos con el extremo reductor libre sobre el ácido murámico.

1.2.1 Estudio de interacción de la N,N',N''triacetilquitotriosa con la WGA

En primer lugar, se estudió la interacción de WGA con quitotriosa, ligando conocido de lectinas con dominio heveína en general [11] y de la WGA en particular [21], como control positivo de la aproximación experimental basada en experimentos de STD. Para ello se preparó una solución stock de quitotriosa 15 mM en agua deuterada.

La quitotriosa, con el extremo reductor libre, está presente en solución acuosa en equilibrio entre ambos anómeros α y β , en proporción 60:40 a favor del anómero α .

Se adquirieron los correspondientes espectros de referencia de una muestra de quitotriosa libre (2mM) para comprobar la no existencia de señales de STD inespecíficas en ausencia de proteína. Los espectros se registraron a dos frecuencias de saturación: 100 ppm (frecuencia *off*-resonance) y 7 ppm (frecuencia *on*-resonance en la zona aromática donde el ligando no presenta ningún protón no intercambiable). Se preparó otra muestra de quitotriosa (1mM) en presencia de WGA (en relación molar

100:1) y se registraron los correspondientes espectros *off*-resonance y *on*-resonance con las mismas frecuencias de irradiación (Figura 3.7).

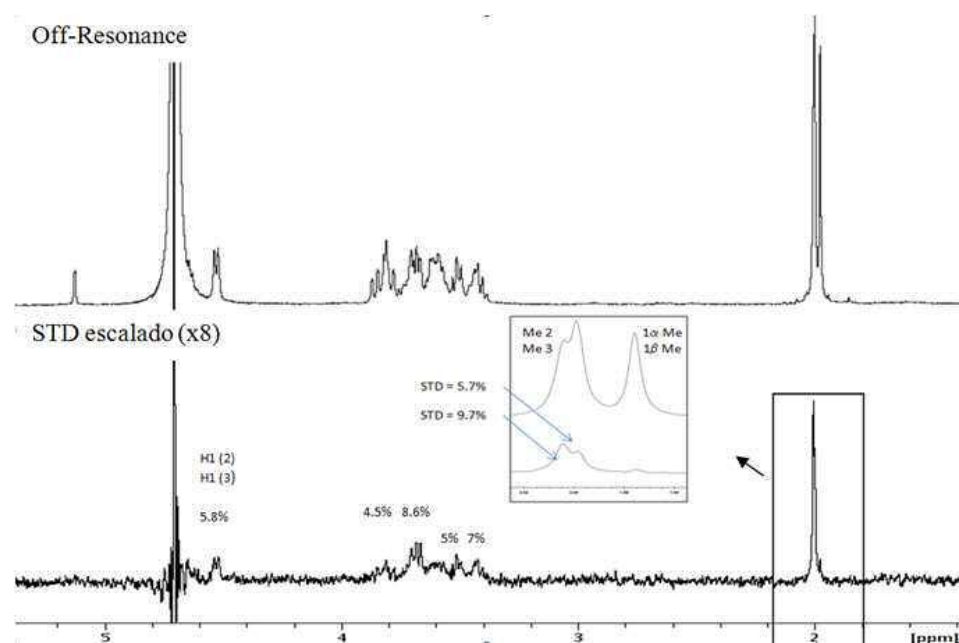


Figura 3.7 ^1H RMN (500MHz a 25°C pH=6 en D_2O) de los experimentos STD para la interacción de la proteína WGA con la quitotriosa.

Analizando los resultados desde el punto de vista del ligando, se observaron cambios en las intensidades de las señales de los metilos cuando se irradiaban los protones de la proteína que tienen un desplazamiento químico en la zona aromática (cadena laterales de los residuos aromáticos). La tabla 3.1 se muestran los valores de STD obtenidos para cada protón, principalmente en la zona de desplazamiento químico de alrededor de 2 ppm, correspondiente a los metilos del triacetilquitotriosa (Figura 3.7).

Tabla 3.1 Desplazamientos químicos de los diferentes protones de la N,N',N'' -Tri-acetilquitotriosa en un espectro ^1H de RMN a 25°C en 500MHz junto con sus valores correspondientes de STD en presencia de la WGA.

Protón (residuo)	Desplazamiento Químico	Intensidad de STD	Porcentaje Normalizado
H1(2) ,H1(3)	4.52 ppm	5.8	60%
H2 α , H2(2), H2(3), H3 α , H5 α , H6	3.75-3.88 ppm	4.5	46%
H2 β , H3 β , H3(2),H6	3.75-3.65 ppm	8.6	89%

H3(3), H4 α , H4 β , H4(2), H5 β , H5(2)	3.55-3.65 ppm	5	52%
H4(3), H5(3)	3.55-3.35 ppm	7	72%
CH ₃ (2), CH ₃ (3)	1.99-2.02 ppm	5.7-9.7	59%-100%
CH ₃ (1) α y β	1.95-1.99 ppm	1	10%

En base a los espectros ROESY y TOCSY del ligando disuelto en agua (con 10% D₂O) adquiridos a 600MHz (Figura 3.8), fue posible diferenciar, gracias a las correlaciones con los HN de los grupos acetamido, las señales correspondientes a los metilos de dichos grupos de los residuos 2 y 3 (1,89-1,92 ppm) de las señales correspondientes al residuo del extremo reductor (tanto del anómero alfa como beta que solapaban a 1,875 ppm). Sin embargo, no fue posible diferenciar inequívocamente las señales de los metilos de los residuos 2 y 3 a pesar de que se diferenciaban ligeramente en el espectro monodimensional (Figura 3.7).

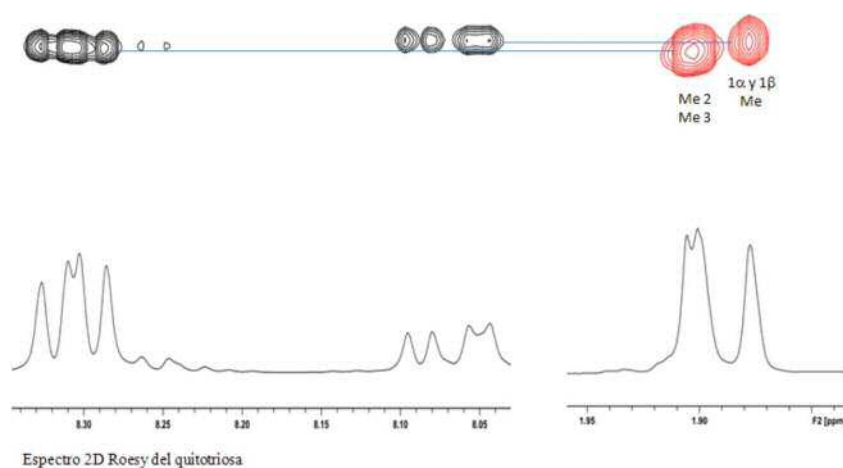


Figura 3.8 Superposición del espectro ¹H RMN de una muestra de quitotriosa en H₂O/10% D₂O con su correspondiente espectro Roesy registrados a 25°C en un campo de 600MHz. En negro se muestran los picos de correlación ROESY HN-CH₃ y en rojo los picos diagonales correspondientes a los grupos metilo.

A partir de los resultados de transferencia de saturación de la WGA a los protones de la quitotriosa se pueden proponer las zonas del ligando que están más cerca de la proteína (*epítipo*) [36]. Los grados de saturación se calculan mediante la resta de las intensidades de las señales en el espectro de *on-resonance* del espectro de referencia *off-resonance* (espectro diferencia, STD). Cada señal del espectro de diferencia se compara con el espectro de referencia *off-resonance* y se normaliza, dando el 100% a la señal que

recibe la máxima saturación. En la figura 3.9 se representan en forma de mapa cualitativo los valores normalizados de la saturación transferida.

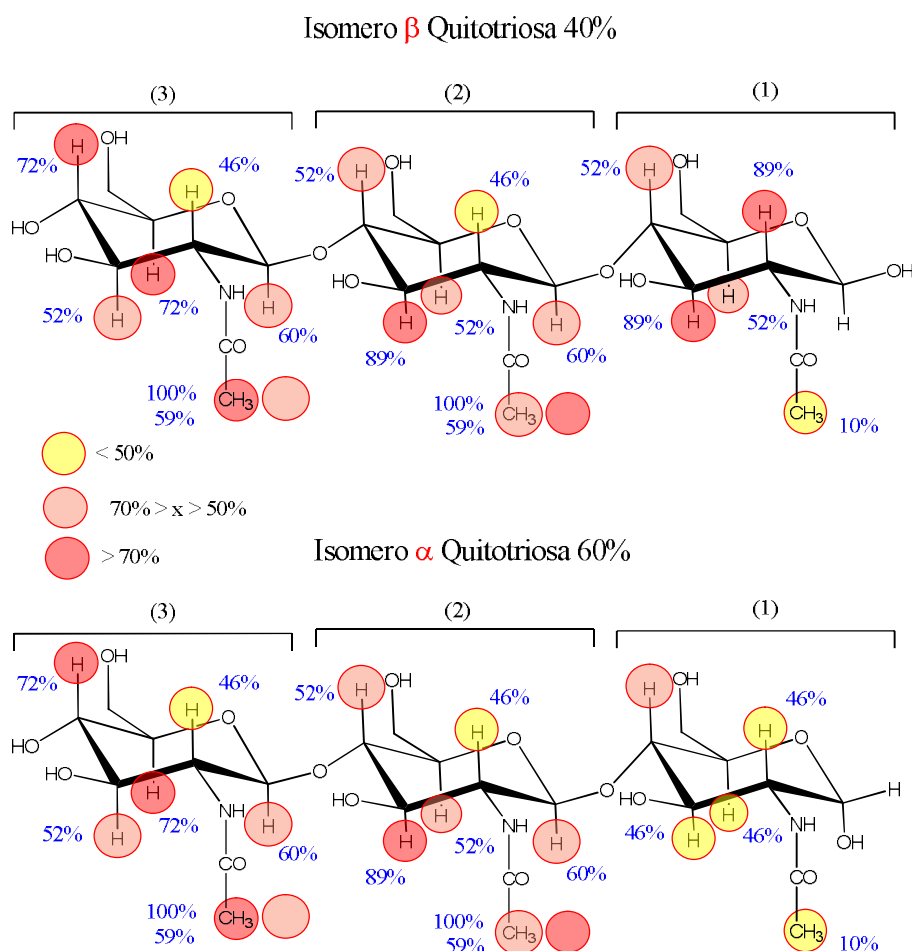


Figura 3.9 Representación del grado relativo de saturación de cada protón de la estructura de la quitotriosa, abajo anómero β , arriba anómero α . El color rojo indica un porcentaje superior o igual al 70%, mientras que el rojo transparente indica un porcentaje entre el 50 y el 70% y el amarillo un porcentaje por debajo del 50%.

El análisis de los epítopos obtenidos de los experimentos de STD de la quitotriosa en presencia de la WGA muestra que la mayoría de los protones de los residuos de GlcNAc 3 y 2 del extremo no reductor de la quitotriosa reciben una saturación significativamente superior al residuo de GlcNAc del extremo reductor. Este hecho se observa inequívocamente en los grupos metilo de los residuos 2 y 3 (entre 6-10%), respecto del metilo del residuo del extremo reductor 1 (1%). De hecho, a la vista de los

dos modos de unión de la figura 3.5 se observa que en el modo A, el metilo del grupo acetamido del residuo 3, situado sobre el subsitio +1, queda encajado dentro del sitio de unión, estableciendo interacciones de tipo CH- π con el residuo aromático en posición 30. En el caso del modo B, es ahora el metilo del residuo 2 el que queda encajado frente al anillo aromático 30. Sin embargo, en ambos modos, el metilo del extremo reductor 1 queda alejado de la superficie de la proteína. Adicionalmente, también se observan diferencias en la distribución de las intensidades relativas de saturación entre los protones de los dos anómeros α y β . El trisacárido con configuración α recibe claramente menos saturación en los protones H-2, H-3 y H-5 del anillo de piranosa que el correspondiente trisacárido en configuración β . Este resultado sugiere que la interacción del residuo de la quitotriosa con la WGA puede ser distinta dependiendo de la conformación del anómero, siendo desfavorecida la configuración α . Por lo general, se había supuesto que los dominios de heveína interaccionan preferentemente con el extremo no reductor de los oligómeros de la quitina [25, 37]. Sin embargo, resultados de nuestro grupo indican que el reconocimiento de la quitina por heveína implica un sitio de unión más extendido en el que oligómeros de más de dos residuos pueden adoptar distintos modos de unión por desplazamiento de las unidades de GlcNAc (Figura 3.5) [5]. Estudios cristalográficos [20, 26-27] y de RMN [5, 11, 29, 38-41] han mostrado que la unidad de GlcNAc del extremo reductor en la posición +2 contribuye a la estabilidad del complejo mediante una interacción CH- π entre el azúcar y el anillo aromático en posición 21 (numeración relativa a la heveína, figura 3.3). Esta interacción se ve favorecida en el caso de la conformación anomérica beta, como se muestra en los estudios de reconocimiento de oligosacáridos cortos (GlcNAc)₂₋₃, tanto libres o con la posición anomérica bloqueada [29, 40], lo que es esperable para el polisacárido de quitina, que tiene todas sus uniones beta. Sin embargo, si el residuo de la posición +2 se presenta en configuración alfa, el hidroxilo de la posición anomérica apunta sobre el anillo aromático de la posición 21 y ejerce un efecto desestabilizador en la interacción.

1.2.2 Estudio de interacción del WGA con el GMDP

A continuación se realizaron los estudios de interacción entre la WGA y el GMDP, como modelo de peptidoglicano. Análogamente a lo que ocurría con la quitotriosa, el GMDP tiene el residuo de ácido murámico con su extremo reductor libre y, por tanto, está presente en disolución en forma de equilibrio anomérico α : β , en proporción 60:40,

en favor del anómero α . La asignación de las señales de resonancia del GMDP se realizó mediante una combinación de experimentos TOCSY, NOESY y ROESY. En la figura 3.10 se representan los espectros monodimensionales en ausencia y en presencia de proteína, con las asignaciones de algunas señales significativas. Los datos se recogen en la tabla 3.2, con los valores de desplazamientos químicos de ^1H RMN, obtenidos a 600 MHz en un tampón deuterado de fosfato sódico, pH 6 a 298K. La comparación de los espectros de ^1H -RMN en ausencia y presencia de proteína (Figura 3.10 espectros A y B respectivamente) puso en manifiesto el ensanchamiento de algunas señales del espectro en presencia de la proteína. Este ensanchamiento es selectivo, especialmente para las señales correspondientes al residuo de ácido murámico en su forma β . Así, por ejemplo, la señal de 4,6 ppm, que corresponde al H1 del resto de murámico β prácticamente desaparece, mientras que la señal del H1 del anómero α (señal por encima de 5 ppm) está claramente definida. Este ensanchamiento de las señales en presencia de la proteína es debido a la variación de la velocidad de relajación transversal. El fenómeno se debe al cambio del tiempo de correlación del ligando al unirse a la proteína y también al proceso de intercambio químico entre los dos estados, asociado y libre [42].

El ensanchamiento observado indica que el GMDP es reconocido por la WGA, hecho que se confirmó con el análisis de los experimentos de transferencia de saturación, STD. Como en el caso de los experimentos con la quitotriosa, los espectros STD se obtuvieron por la diferencia de los espectros registrados a dos frecuencias de saturación: 100 ppm (frecuencia *off*-resonance) y 7 ppm (frecuencia *on*-resonance) y confirmaron inequívocamente, después de comparar las intensidades entre el espectro de referencia (*off*-resonance) y el espectro de STD, el fenómeno de reconocimiento del ligando GMDP por la parte de la WGA (Figura 3.10, espectros B y C respectivamente). Adicionalmente y análogamente a lo observado con el ensanchamiento de líneas, la transferencia de saturación fue selectiva para algunos protones del ligando. Este hecho permitió definir el epítipo de unión. Además, también se observó selectividad entre las dos formas anoméricas en equilibrio, lo que permitió determinar cuál de los dos anómeros era seleccionado preferentemente por la WGA.

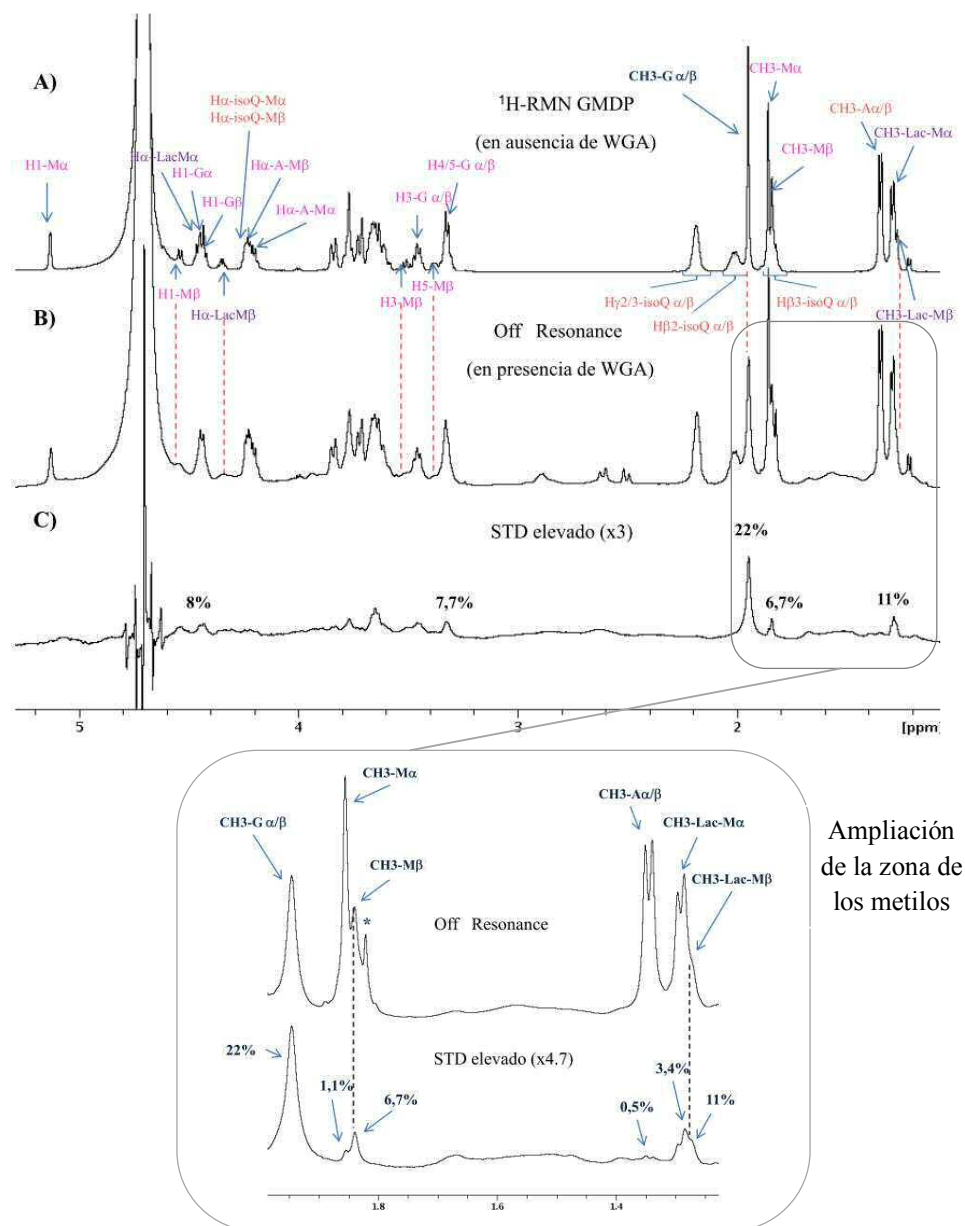


Figura 3.10 A) Espectro ^1H -RMN del GMDP en ausencia de proteína a 298K en D_2O a pH 6 a 600 MHz. B) Espectro *off* resonance del experimento STD del GMDP (2 mM) en presencia de 0,2 mM WGA (tiempo de saturación = 2s; frecuencia *off*-resonance: 100 ppm; frecuencia *on*-resonance: 7 ppm) las líneas discontinuas que conectan los espectros sin y con proteína indican picos ensanchados. C) Espectro de diferencia de STD. * Impureza en la muestra (El espectro está referenciado a la señal de H_2O a 4,7 ppm).

Los resultados de transferencia de saturación se recogen en la tabla 3.2. La señal del metilo del resto de GlcNAc (que no se diferencia para las dos formas en equilibrio del residuo de murámico) es la que recibe la máxima saturación, con un valor de STD del

22 %. Este resultado está de acuerdo con lo observado para la quitotriosa. De hecho, es esperable según lo descrito para el modo de unión de quitobiosa y quitotriosa a diversos dominios heveína [29-30, 43], en los que el grupo acetamido de la GlcNAc está en estrecho contacto con residuos de la proteína (ver figura 3.4). El resto de los protones de la parte del carbohidrato tienen valores de intensidades de STD entre 2 y 12 %, mientras que los valores de saturación correspondientes a los protones de la región peptídica no alcanzan valores superiores al 4%, indicando que esta parte establece pocos contactos con la proteína (tabla 3.2).

A partir de los resultados de la tabla 3.2 se puede distinguir, de manera clara, los protones que están en estrecho contacto con la proteína permitiendo dibujar el mapa del epítipo de unión en el ligando (Figura 3.11).

Tabla 3.2 Desplazamiento químico de los diferentes protones del GMDP a 298K en 600MHz junto con sus valores correspondientes de STD en presencia de la WGA.

Protón (residuo)	Desplazamiento Químico	Intensidad de STD	Porcentaje Normalizado
CH3-Lac-Mβ	1.388	11%	49%
CH3-Lac-Mα	1.406	3.5%	15.8%
CH3-Alanina α/β	1.462	0.5%	2.3%
CH3-Mβ	1.956	6.7%	30.4%
CH3-Mα	1.973	1.1%	5.1%
HB3-Isoglutamina α/β	1.964	< 1%	----
CH3-GlcNAc α/β	2.064	22%	100%
HB2-Isoglutamina α/β	2.12	< 1%	----
HG2/3-Isoglutamina α/β	2.301	< 1%	----
H4/H5-Gα/β	3.40 - 3.47	7.7%	34.9%
H5-M β	3.495	Ensanchamiento	----
H3-G α/β	3.54 - 3.56	10.5%	47.6%
H3-M β	3.62	(18%) Ensanchamiento	----
H2-M β / H3-M β	3.72	8.4%	38.1%
H2-G α/β	3.75	10.6%	48.1%
H6b-G α/β	3.76 - 3.81	12.5%	56.71%
H2-Mα / H4-Mα	3.82	2.2%	10.2%
H4-M β	3.85	Ensanchamiento/Solapamiento	----
H6 a/b-M α/β	3.86 - 3.92	6%	27.2%
H6a-G α/β	3.79 - 3.96	4%	17.7%
Hα-Alanina-M β	4.316 (4.29-4.35)	< 4% Solapamiento	----
Hα-Alanina-M α	4.33 (4.29-4.35)	< 4% Solapamiento	----
Hα-Isoglutamina M α/β	4.34 - 4.38	< 4% Solapamiento	----
Hα-Lac M β	4.462	Ensanchamiento	----
H1-G β	4.54 (4.52-4.55)	(13%) Solapamiento	(58%)
H1-G α	4.554 (4.54-4.57)	(8%) Solapamiento	(36%)
Hα-Lac M α	4.57 (4.52-4.60)	(6%) Solapamiento	(25%)
H1-M β	4.655	Ensanchamiento	----
H1-M α	5.247	2%	8.6%

Los valores se dan en ppm con referencia al metilo del TSP a 0 ppm. En el caso de las señales ensanchadas y/o solapadas los valores de saturación son aproximados.

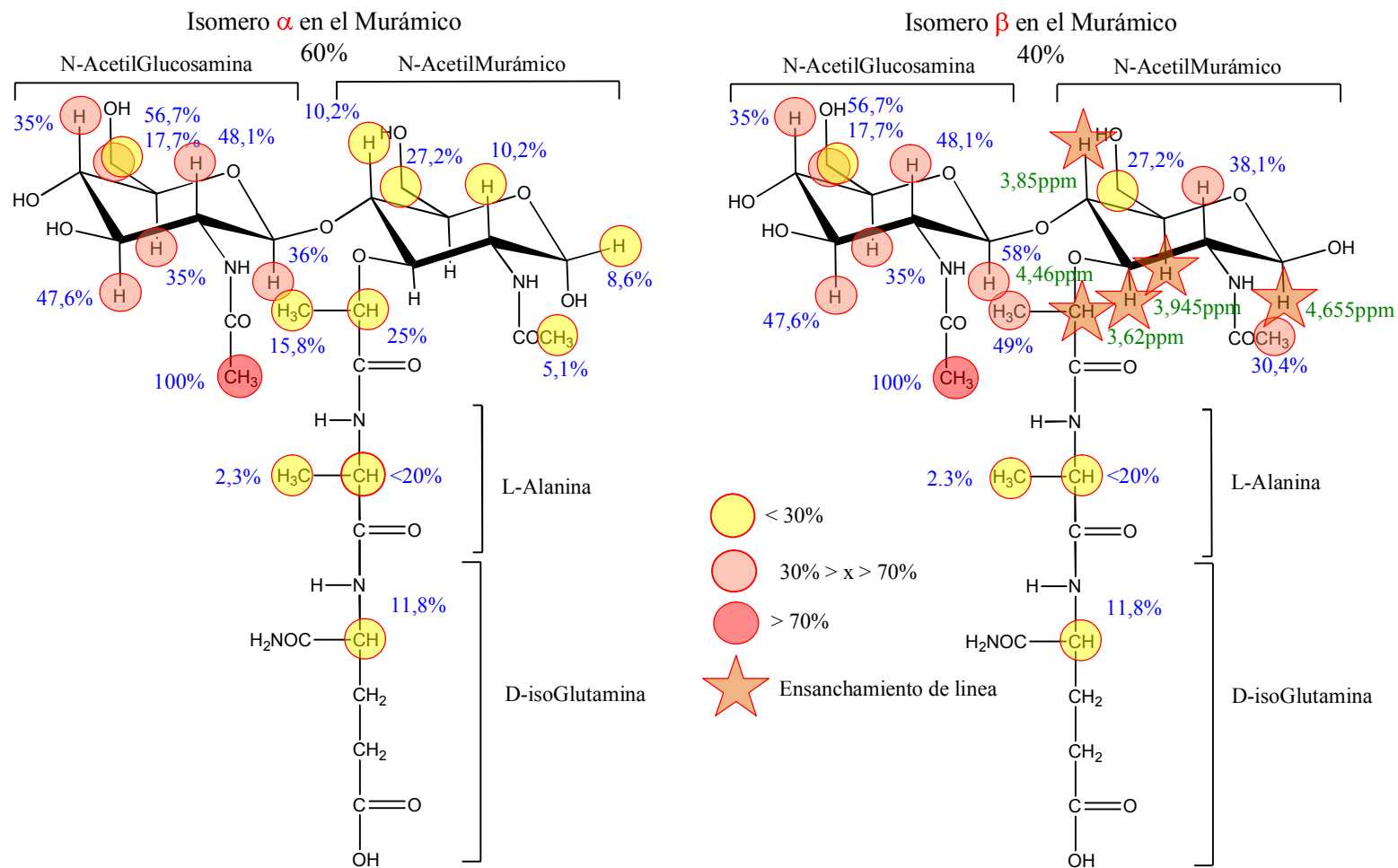


Figura 3.11 Estructura del compuesto GMDP (N-Acetyl-l-D-glucosaminy-l-N-acetylmuramyl-L-alanyl-D-isoglutamine) y los valores relativos de STD para cada uno de los protones en presencia de WGA.

A la vista la tabla 3.2 y la figura 3.11 queda claro que la unión al GMDP se establece principalmente desde la porción de la N-acetilglucosamina. De hecho, los protones del grupo metilo de su resto acetamido son los que reciben la máxima saturación en toda la molécula del ligando. Por el lado, para el ácido N-Acilmurámico, los porcentajes de saturación obtenidos para los protones de cada una de las dos configuraciones anoméricas α y β , muestran que tanto la señal del metilo del ácido láctico como la señal de metilo del grupo acetamido del N-acetilmurámico alcanzan valores superiores de saturación cuando el ligando presenta la configuración β (ver el zoom del espectro de la figura 3.10).

Con objeto de obtener un modelo 3D de la estructura del complejo GMDP-WGA, se realizó una superposición del GMDP, tomado de la estructura cristalina de un complejo de GlcNAc-MurNAc-L-Ala-D-Glu (el GMDP con ácido glutámico en vez de isoglutamina) con una transglucosilasa lítica de *E. coli* (código pdb: 1d0k) [44], sobre las coordenadas de la N-acetilglucosamina presente en el sitio de unión del dominio B de la WGA de la estructura cristalina de un complejo de WGA con N-Acetilglucosamina (código pdb: 2UVO) [27] (Figura 3.12).

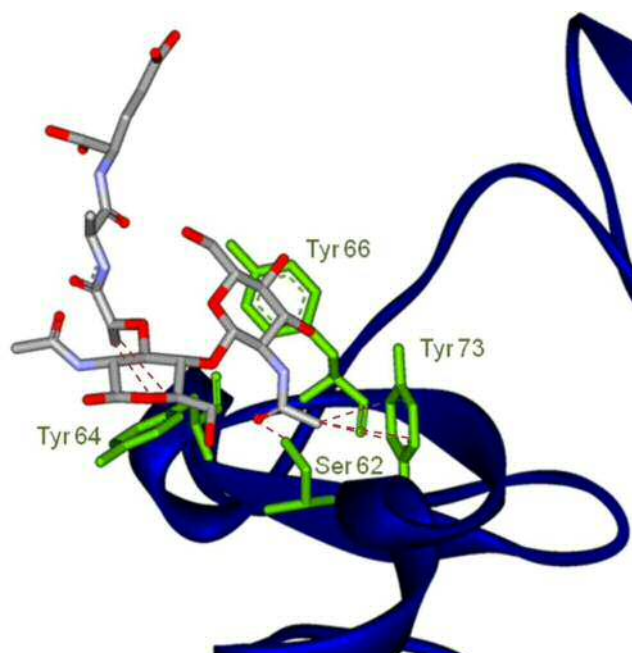


Figura 3.12 "Docking" manual del β -GMDP (a partir de la estructura presente en PDB 1D0K) sobre el dominio B de la WGA (color azul) (PDB 2UVO). En verde, aminoácidos involucrados en la interacción..

Como se puede observar, el resto de GlcNAc del GMDP encaja en el sitio de unión, como es esperable para un dominio heveína. El residuo de Ser 62 establece un enlace de hidrógeno con el carbonilo del acetamido; el residuo Tyr 73 establece una interacción de van der Waals con el grupo metilo del mismo residuo. Este hecho está en concordancia con el epítipo descrito anteriormente, en el que este grupo metilo (CH₃-GlcNAc α/β tabla 3.2) recibe la máxima saturación (22%). Análogamente, los residuos aromáticos Tyr 66 y Tyr 64 establecen interacciones de apilamiento con las dos unidades de GlcNAc y MurNAc consecutivas [45]. Este modelo justifica las saturaciones observadas (> 7% en casi todos los protones de los anillos de piranosa) y las perturbaciones de la forma de línea de los protones de ambos residuos. La parte peptídica del ligando en este modelo queda esencialmente expuesta al disolvente, de acuerdo con los bajos índices de saturación observados para los residuos de alanina y isoglutamina (< 4%, tabla 3.2). El grupo metilo del resto de ácido láctico en la posición 3 del ácido murámico (CH₃-Lac-M β , tabla 3.2) apunta hacia la Tyr 64 y la superficie de la proteína, justificándose su significativa saturación (11%).

Alternativamente, al superponer el MurNAc sobre el GlcNAc presente en la estructura cristalina (pdb: 2UVO) para situarlo en el sitio +1 (considerando el segundo modo de unión mostrado en la figura 3.5-B), el residuo GlcNAc del extremo no reductor del GMDP quedó en la posición -1). En esta disposición, la parte peptídica de la posición 3 del resto de MurNAc mostró severos choques estéricos con la proteína. Es decir, el modelo estructural presentado en la figura 3.12, junto con los resultados del STD permiten proponer que el extremo no reductor del ligando (GlcNAc) está bien acomodado en el sitio de unión de la proteína (estableciendo una interacción tipo CH- π de apilamiento carbohidrato-aromático sobre el anillo aromático de la tirosina 66 y posicionando el grupo acetamido cerca de la tirosina 73), mientras que el extremo reductor proporciona contactos menores con la proteína. Cuando se cambia la configuración de la posición anomérica de β a α del resto de murámico sobre el modelo, el hidroxilo del extremo reductor adopta una configuración axial que apunta hacia la Tyr 64. Esta interacción es desfavorable para la unión y está en concordancia con los datos experimentales, que indican una clara preferencia por el anómero β .

1.3 CONCLUSIONES

De los resultados obtenidos en este capítulo, se pueden extraer las siguientes conclusiones:

- La unión del triacetilquitotriosa al WGA se establece principalmente a través del extremo no reductor de la quitotriosa ya que la mayoría de los protones de los residuos de GlcNAc 3 y 2 reciben una saturación significativamente superior al residuo de GlcNAc del extremo reductor.
- La triacetilquitotriosa con configuración α recibe claramente menos saturación en los protones H-2, H-3 y H-5 del anillo de piranosa reductor que el correspondiente trisacárido en configuración β .
- La unión de la WGA al GMDP, relacionado con el peptidoglicano se confirmó inequívocamente, después de observar ensanchamientos selectivos en algunas señales del ligando y por experimentos de STD después de comparar las intensidades entre el espectro de referencia (*off-resonance*) y el espectro de STD.
- La unión al GMDP se establece principalmente desde la porción de la GlcNAc que se justifica porque su grupo acetamido recibe la máxima saturación observada y en el resto de los protones del anillo de piranosa, las saturaciones son superiores a las correspondientes observadas en el resto del N-acetilmuramico.
- La unión del GMDP a la WGA usando como modelo el dominio B (el segundo dominio tipo heveína en la secuencia de la WGA) se establece según el modo de unión donde el anillo de GlcNAc se sitúa sobre la Tyr 66, donde la parte peptídica del GMDP queda esencialmente expuesta al disolvente, de acuerdo con los bajos índices de saturación observados para los residuos de alanina y isoglutamina (< 4%). El grupo metilo del resto de ácido láctico en la posición 3 del ácido murámico (CH₃-Lac-M β , tabla 3.2) apunta hacia la Tyr 64 y la superficie de la proteína, justificándose su significativa saturación (11%).

2 DOMINIO LYSM

2.1 INTRODUCCIÓN Y ANTECEDENTES

En la naturaleza, las plantas están constantemente expuestas a una amplia gama de microorganismos presentes en su entorno. La interacción con algunos puede ser perjudicial, como en el caso del parasitismo u otros procesos infectivos [46], pero también puede ser beneficiosa, como en el caso de la simbiosis. Con el fin de distinguir entre microorganismos patógenos y beneficiosos, las plantas presentan en sus membranas una serie de proteínas receptoras implicadas en el reconocimiento de patrones de carbohidratos comúnmente asociados con las superficies microbianas, tales como, la quitina de hongos [47-51] y el peptidoglicano de bacterias [46, 52-53]. Entre ellos se destacan los receptores quinasas tipo “*Lysin Motive Receptor-Like Kinase*” (LysM-RLK). Estos receptores contienen los denominados dominios LysM de aproximadamente 50 aa en la región extracelular [7, 9, 54-55] y un dominio intracelular con actividad quinasa que media las etapas de reconocimiento e infección o colonización/simbiosis [56-57]. Se encuentran distribuidos en virus, bacterias, hongos, plantas y animales que dependiendo de su localización, pueden estar implicados en funciones líticas, inmunológicas, en la infección, en la morfogénesis y también en la simbiosis bacteria-planta leguminosa [7, 47, 49, 51, 58-63].

Los dominios LysM fueron descritos originalmente en la lisozima del fago $\Phi 29$ de *Bacillus* [64-65], que degrada las paredes celulares bacterianas por hidrólisis de los enlaces glicosídicos entre MurNAc y GlcNAc del peptidoglicano [7, 66]. Más tarde se identificaron en muchas enzimas bacterianas y fúngicas implicadas en la degradación de la pared celular [7, 53-55] como las quitinasas, enzimas que descomponen los polímeros de GlcNAc, [7], así como en diversas hidrolasas, implicadas en la remodelación del peptidoglicano de la pared celular durante la división bacteriana [67-68]. Además, se ha descrito la participación del dominio LysM en otro tipo de funciones, como en la morfogénesis asociada con el montaje y el desarrollo de la capa de endosporas [69] y en el proceso de infección de plantas por *Cladosporium*. En este proceso está involucrada una proteína (Ecp6) que consigue bloquear la defensa inmune de plantas uniéndose a los fragmentos de quitina y permitiendo al hongo infectar a la planta, sin ser detectado [70]. Por otro lado y con una estrategia opuesta, desde la perspectiva de las plantas, que a través de sus correspondientes dominios LysM, inician el proceso de simbiosis en la

raíces con las bacterias beneficiosas mediante el reconocimiento de los factores de nodulación (moléculas relacionadas con la quitina) [9, 71-72] (Figura 3.13).

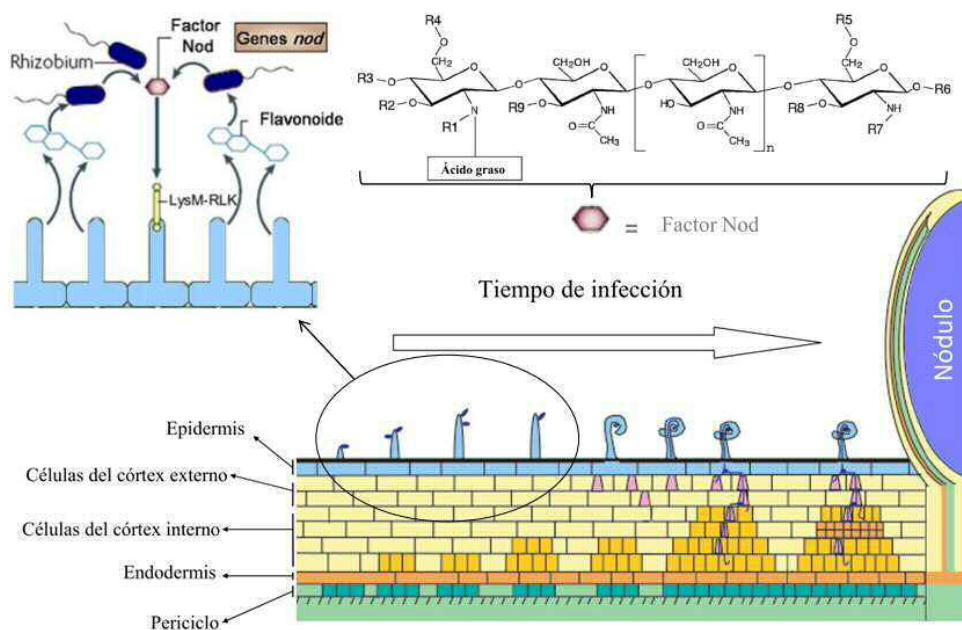


Figura 3.13 Representación esquemática del proceso de infección y desarrollo de un nódulo, haciendo enfoque sobre la inducción de los genes Nod en las bacterias que, tras la percepción de los flavonoides derivados de la planta, permitirán a las bacterias la liberación de los factores de nodulación (lipoquitooligosacáridos LCO) reconocidos por los dominio LysM-RLK en la superficie de las células de las raíces de la planta.

Brevemente, la simbiosis comienza con un intercambio de señales moleculares específicas entre la planta hospedadora y la bacteria [9, 71, 73-77]. La planta leguminosa produce, a través de las células de la raíz, un cóctel de moléculas formadas por flavonoides capaces de atraer a las bacterias e inducir en ellas una serie de regulaciones que les llevan a secretar los lipoquitooligosacáridos (LCO) que actúan como factores de nodulación (Nod Factor). Estos factores son reconocidos por los receptores LysM-RLK expresados en la superficie de las células de las raíces. Con esta interacción se disparan en la raíz de la planta los procesos que permiten primero la infección o entrada de la bacteria simbiote [77-79], y posteriormente la formación del nódulo especializado en la fijación del nitrógeno [80].

En este sentido, el grupo de la Dra Cullimore del instituto INRA de Toulouse (Francia) había descrito varias proteínas del tipo LysM-RLK en el trébol *Medicago truncatula* como posibles candidatas para actuar de receptoras de los lipoquitooligosacáridos secretados por la bacteria rizobium como factores de nodulación [9, 81]. Entre estas proteínas cabe mencionar las derivadas de los genes NFP y LyK3 que, en ambos casos, cuentan con tres dominios LysM consecutivos y conectados entre sí por pequeños segmentos peptídicos. Se ha propuesto que los residuos de cisteína presentes en cada dominio LysM ayudan a mantener el correcto plegamiento.

Cuando se empezó trabajar sobre el dominio LysM en el desarrollo de esta Tesis, la base de datos de proteínas (Protein Data Bank) sólo contaba con la presencia de tres estructuras 3D de este tipo de dominios. Dos eran de origen bacteriano, el segundo dominio LysM de la proteína lítica de membrana (MltD) de *E. Coli*, determinado mediante RMN (código pdb: 1E0G) [55] y un dominio de la proteína ykuD de *Bacillus subtilis* (código pdb: 1Y7M) cristalizada a una resolución de 2.0 Å [82]. La tercera estructura depositada pertenecía a una proteína hipotética humana SB145 (código pdb: 2DJP), también determinada por RMN. Todos estos pdb's, desde el punto de vista estructural, coincidían a que el dominio LysM mostraba una estructura tipo $\beta\alpha\alpha\beta$, con dos α -hélices empaquetadas en el mismo lado de la hoja β antiparalela [55].

En los últimos años se han sumado a la base de datos otras cuatro nuevas estructuras de dominios LysM. Dos fueron obtenidas mediante RMN. En una de ellas, el dominio LysM forma parte de la enzima GpX del bacteriófago P2 y tiene una función general de unión a peptidoglicano de la membrana para facilitar la inyección del ADN (código pdb: 2LTF) [83]. El otro dominio LysM forma parte de una proteína de la planta de arroz que se expresa en respuesta a la infección por hongos (código pdb: 2L9Y) [84]. Otras dos estructuras fueron determinadas mediante cristalografía. Una corresponde al receptor de quitina AtCERK1 en *Arabidopsis*. Está compuesto por 3 dominios LysM e implicado en la inducción de la respuesta inmune (códigos pdb : 4EBY y 4EBZ, en este caso en presencia de un tetrasacárido ligando) [62]. La otra corresponde a la proteína Ecp6 del hongo *Cladosporium* mencionada arriba, que también contiene tres dominios LysM (códigos pdb 4B9H y 4B8V en este caso también en presencia de quitooligosacáridos) [70] (Figura 3.14).

Dada la importancia de este dominio, en el presente proyecto de investigación nos propusimos estudiar la interacción carbohidrato-proteína en los procesos de reconocimiento de los factores de nodulación, y en los que se ha propuesto que participan dominios LysM. En una primera etapa se eligió el dominio LysM de la transglicosidasa lítica de *E. coli* como un sistema modelo para el estudio de este tipo de interacciones usando como ligandos estructuras poliméricas como la quitina, el peptidoglicano crudo o fragmentos de estas estructuras como el GMDP y un quitopentaoligosacárido, como modelos más sencillos. En una siguiente etapa, se pretendía estudiar el dominio LysM de proteínas receptoras de factores de nodulación del sistema de simbiosis bacteriana en plantas. En particular, el segundo dominio LysM presente en la parte extracelular de la proteína Lyk-3 que se había propuesto como receptor del tipo “LysM-RLK” de los factores de nodulación de tipo LCO, en procesos de simbiosis en *Medicago truncatula* [85].

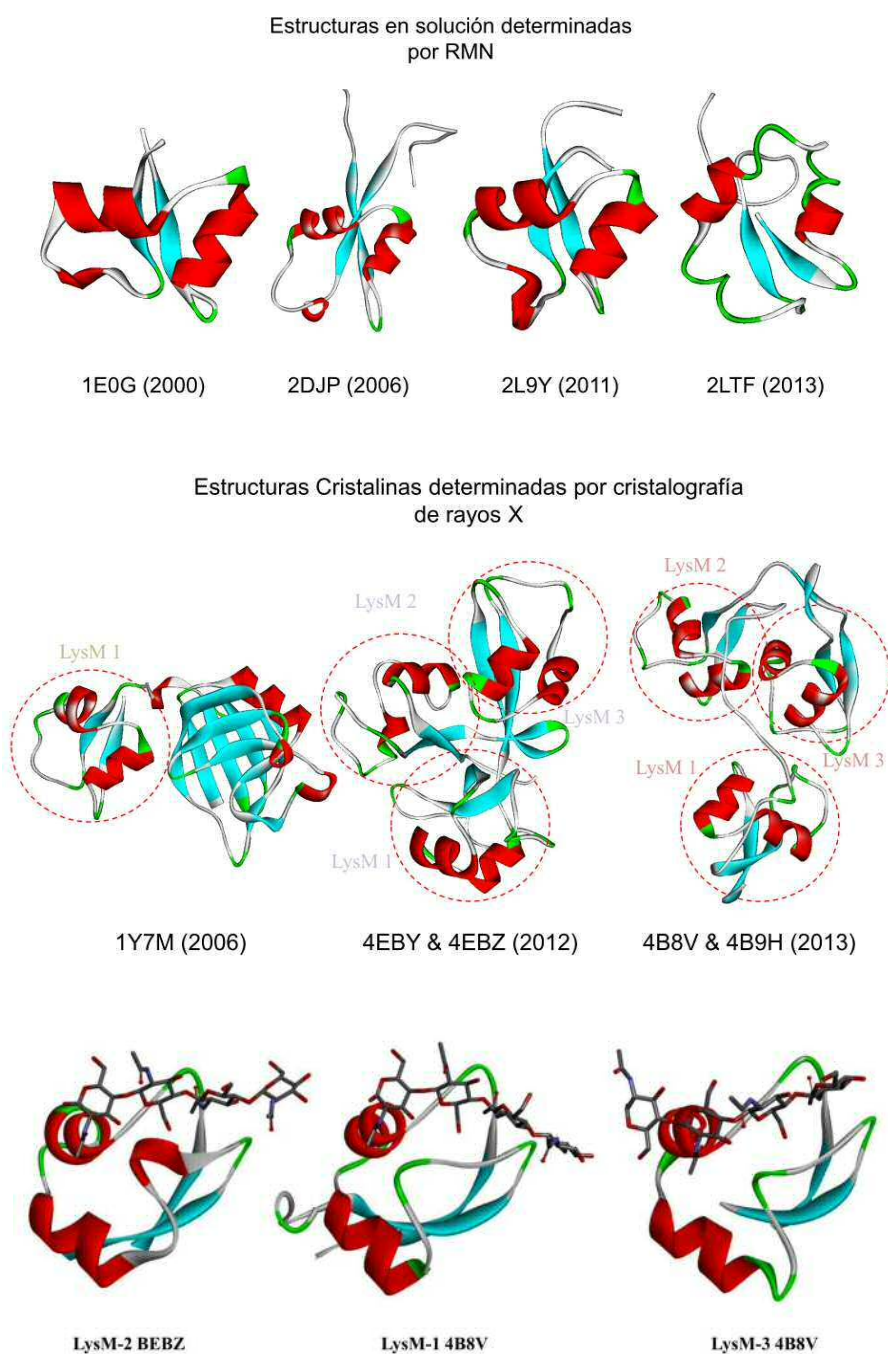


Figura 3.14 Representación de las estructuras conocidas de los dominios LysM determinadas tanto en solución por RMN (panel superior) como por cristalografía de rayos X (panel medio) con sus códigos pdb correspondientes. En el panel inferior, se representan los dominios LysM cristalizados con ligandos oligosacáridicos.

2.2 RESULTADOS Y DISCUSIÓN

2.2.1 Secuencias de dominios LysM

Segundo dominio LysM de la transglucosidasa lítica de *E.coli*

Para estudiar el segundo dominio LysM de la transglucosidasa lítica de *E. coli* dispusimos de los plásmidos clonados con la secuencia entre los aminoácidos 389 y 452 (GenBank número de acceso AAC73316) que contienen introducidas dos cisteínas en distintas posiciones relativas respecto a los extremos N- y C-terminal para estabilizar el plegamiento [86]. Las secuencias codifican para los dominios LysM-CAM-1, LysM-CAM-2 y LysM-CAM3 y tienen un segmento adicional de 17 aminoácidos que incluye polihistidinas de 6 residuos y un sitio de proteólisis sensible a trombina (L-V-P-R-G-S) (tabla 3.3).

Tabla 3.3 Secuencias de los tres mutantes del segundo dominio LysM de la transglucosidasa lítica de *E. coli*.

Dominio y sus mutantes	Secuencia	Mutaciones (descripción de cambios)
LysM-2 MltD Original	AQRLANNSDSITYRVRKGDSLSSIAKRH GVNIKDVMRWNSDTANLQPGDKLTLF VKNNMPDS	Secuencia de referencia LysM de 389–452 MltD <i>Escherichia coli</i>
LysM-CAM-1	MRGSHHHHHHGLVPRGSAQRLANNCD SITYRVRKGDSLSSIAKRHGVNIKDVMR WNSDTANLQPGDKLTLFVKCNNMPDS	Mutaciones S396C y N445C en posiciones -1 y 49 relativas al dominio LysM.
LysM-CAM-2	MRGSHHHHHHGLVPRGSAQRLANCSD SITYRVRKGDSLSSIAKRHGVNIKDVMR WNSDTANLQPGDKLTLFVKNCNMPDS	Mutaciones N395C y N446C en posiciones -2 y 50 relativas al dominio LysM.
LysM-CAM-3	MRGSHHHHHHGLVPRGSAQRLACNSD SITYRVRKGDSLSSIAKRHGVNIKDVMR WNSDTANLQPGDKLTLFVKNNCMPDS	Mutaciones N394C y N447C en posiciones -3 y 51 relativas al dominio LysM.

En violeta secuencia de 48 aminoácidos del segundo dominio LysM-MltD. **En azul** extremos N- y C-terminal contiguos al dominio LysM-MltD. *En cursiva*, posiciones sustituidas por cisteína. **En rojo** posición de la cisteína introducida y **en negro**, segmento adicional incluyendo polihistidina y secuencia de corte por trombina.

Dominios LysM de la proteína RLK Lyk-3 de *Medicago Truncatula*

Basándonos en la estrategia descrita por Clarke y col. y en colaboración con la Dra J. Cullimore (INRA, Toulouse (Francia)) se diseñaron una serie de secuencias modificadas

para el segundo dominio LysM de la proteína Lyk3 [85] (51 aa, G99-T149 de la secuencia GenBank: AAQ73159) con objeto de conectar el N-terminal con el C-terminal mediante un puente disulfuro estabilizante que permitiera realizar los estudios en disolución por RMN. Así, se utilizaron las propias cisteínas existentes en la secuencia de Lyk3 o se usaron modificaciones similares a las aplicadas en la preparación de las variantes CAM-1, -2 y -3. Con esta idea, en primer lugar, se utilizaron los extremos de la secuencia del LysM-CAM-1 (tabla 3.3) en las posiciones de las cisteínas introducidas para preparar la construcción Lyk-3-LysM-2-1, insertando entre estos segmentos los 51 aa del dominio LysM.

Alternativamente, se incluyeron los propios extremos N- y C- de Lyk-3-LysM-2 (C90-S157), manteniendo solamente una de las dos cisteínas presentes en cada extremo. Así se preparó la construcción Lyk-3-LysM-2-2. Seguidamente, se diseñaron otras 5 construcciones, manteniendo parte de las secuencias de los extremos N- y C- propias del Lyk3, pero variando las posiciones relativas de las cisteínas respecto del aminoácido inicial del dominio LysM. En estas últimas construcciones, las cisteínas originales presentes en Lyk3 se sustituyeron por alaninas. Los correspondientes plásmidos para estas construcciones fueron realizados por el grupo de la Dra Julie Cullimore (tabla 3.4)

Tabla 3.4 Secuencias de los siete construcciones mutantes del segundo dominio LysM de la proteína RLK Lyk-3 de *Medicago Truncatula*.

Proteína	Secuencia	Mutaciones (descripción de cambios)
Lyk3-LysM-2 Original	NVPFPCECIGGEFLGHVFEYTTKEGDDYD LIANTYYASLTTVELLKKFNSYDPNHIPVK AKINVTVICS CGNS QISKDYGL	Secuencia de referencia Lyk3-LysM-2 N85-L165 Dominio LysM 51 aa, G99-T149 GenBank: AAQ73159
Lyk3-LysM-2-1	MRGSHHHHHHGLVPRGSAQRLANNC GHV FEYTTKEGDDYDLIANTYYASLTTVELLKK FNSYDPNHIPVKAKINVT CNNMPDS	Lyk3-LysM-2 G99-T149 insertada entre las secuencias de los extremos de CAM-1 C en posiciones relativas al dominio LysM: -1 y 52
Lyk3-LysM-2-2	MRGSHHHHHHGLVPRGSVT A ECIGGEFLG HVFYTTKEGDDYDLIANTYYASLTTVELL KKFNSYDPNHIPVKAKINVT VICS A GNS	Lyk3-LysM-2 C90-S157 mutaciones C90A y C154A C en posiciones relativas al dominio LysM: -7 y 54.
Lyk3-LysM-2-3	MRGSHHHHHHGLVPRGSIGGEFC GHVFEY TTKEGDDYDLIANTYYASLTTVELLKKFNS YDPNHIPVKAKINVT VICS A GNS	Lyk3-LysM-2 I93-S157 mutaciones L98C y C154A. C en posiciones relativas al dominio LysM: -1 y 54.
Lyk3-LysM-2-4	MRGSHHHHHHGLVPRGSIGGEFC GHVFEY TTKEGDDYDLIANTYYASLTTVELLKKFNS YDPNHIPVKAKINVT VC A S A GNS	Lyk3-LysM-2 I93-S157 mutaciones L98C, I151C, C152A C154A C en posiciones relativas al dominio LysM: -1 y 53.
Lyk3-LysM-2-5	MRGSHHHHHHGLVPRGSIGGEFC GHVFEY TTKEGDDYDLIANTYYASLTTVELLKKFNS YDPNHIPVKAKINVT CI A S A GNS	Lyk3-LysM-2 I93-S157 mutaciones L98C, V150C, C152A C154A C en posiciones relativas al dominio LysM: -1 y 52.
Lyk3-LysM-2-6	MRGSHHHHHHGLVPRGSIGGE CL GHVFEY TTKEGDDYDLIANTYYASLTTVELLKKFNS YDPNHIPVKAKINVT VC A S A GNS	Lyk3-LysM-2 I93-S157 mutaciones F97C, I151C, C152A C154A C en posiciones relativas al dominio LysM: -2 y 53.
Lyk3-LysM-2-7	MRGSHHHHHHGLVPRGSIGGC FL GHVFEY TTKEGDDYDLIANTYYASLTTVELLKKFNS YDPNHIPVKAKINVT VICS A GNS	Lyk3-LysM-2 I93-S157 mutaciones E96C, C154A C en posiciones relativas al dominio LysM: -3 y 54.

En marrón secuencia del dominio LysM-2-Lyk-3 (51 aa). **En rojo** sustitución de un aminoácido por cisteína. **En azul** secuencias de los extremos de LysM-MtD. **En verde** secuencia original de los segmentos contiguos al dominio LysM-2-Lyk3. **Resaltado en amarillo** sustitución de cisteínas originales por alanina y **en negro**, segmento adicional incluyendo polihistidina y secuencia de corte por trombina.

2.2.2 Expresión & purificación de dominios LysM

2.2.2.1 Construcciones derivadas de MltD de *E. coli* (CAM-1-3)

Inicialmente, las construcciones LysM-CAM-1, 2 y 3 de la tabla 3.3 cedidas por la Dra Clarke se clonaron en células BL21 de *E. Coli*. Desgraciadamente, la expresión de las proteínas en esta cepa resultó ser muy poco eficiente. Alternativamente, se usaron células C41 (DE3) derivadas de cepas de *E. coli* BL21 DE3, recomendadas para el estudio de proteínas tóxicas de membrana [87].

Las células C41 (DE3) se transformaron con los correspondientes plásmidos recombinantes pRSETA, que contenían las construcciones LysM-CAM-1, 2 y 3 (Figura 3.15) (tabla 3.3) y se determinaron empíricamente las condiciones óptimas de expresión en medio TYx2, variando las condiciones de inducción y analizando los extractos crudos en geles de SDS-PAGE, buscando la presencia de nuevas bandas a la altura del marcador de 9 kDa que no aparecían en los extractos de cultivos no inducidos. De este modo, se establecieron las mejores condiciones de expresión de las construcciones correspondientes a los dominios LysM-CAM-1, LysM-CAM-2 y LysM-CAM-3 que correspondían a una inducción con IPTG 1mM en la fase de crecimiento a una D.O.₆₀₀ equivalente a 0,6 a 25°C, seguido por una incubación inmediata durante 20 horas en agitación orbital. (Figura 3.16).

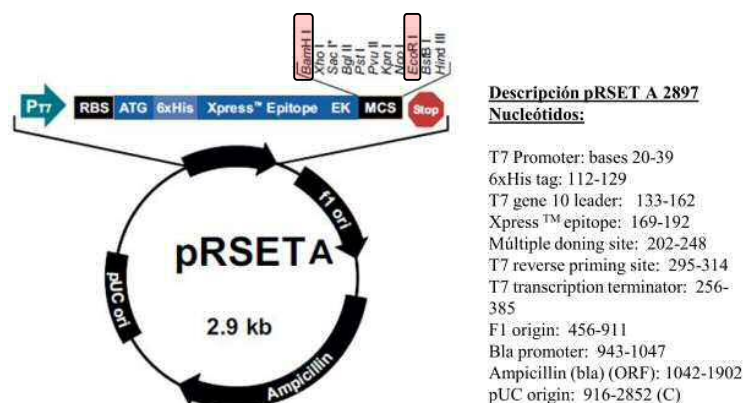


Figura 3.15 Mapa que muestra las principales características del plásmido pRSET A. Las construcciones LysM-CAM se introdujeron entre los sitios de inserción EcoR1 y BamH1 recuadrados.

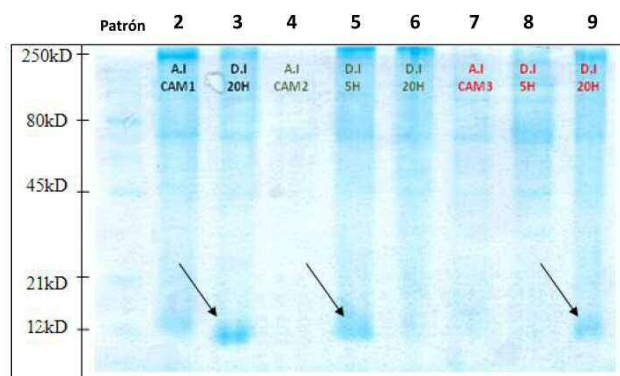


Figura 3.16 SDS-PAGE al 15%. Análisis de extractos crudos de alícuotas de minicultivos para LysM-CAM1, CAM2 y CAM3: Las flechas indican las bandas más intensas después de inducir (~9 kDa). A.I: Antes de inducir, D.I: después de inducir.

La producción de los péptidos recombinantes con la extensión de histidinas permitió el uso de la cromatografía de afinidad en columna de Ni-NTA como primera etapa para su purificación.

Tras optimizar las condiciones de expresión, se logro sobreexpresar y purificar el péptido recombinante LysM-CAM-1 en cultivos escalados a 400-800 ml de forma soluble en las condiciones anteriormente determinadas (Figura 3.17 A). Por el contrario, no se consiguieron rendimientos de expresión aceptables para los péptidos LysM-CAM-2 y LysM-CAM-3 en cultivos a gran escala (Figura 3.17 B)

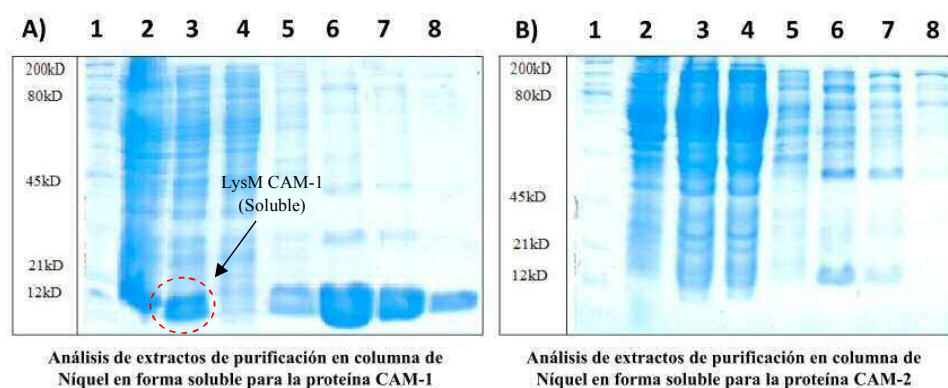


Figura 3.17 SDS-PAGE al 15% de purificación del A) LysM-CAM-1 B) LysM-CAM-2. 1: patrón, 2: antes de inducir, 3: después de inducir, 4: lavado del extracto depositado en la columna de afinidad de Ni²⁺. 5-8 Eluciones de la columna.

En el caso de LysM-CAM-1, la proteína no se aparecía pura en los eluidos con imidazol (a 250 mM) observando contaminantes de alta mas molecular. Por este motivo, se aplico en una columna Superdex tipo Hiloal 16/60 y 26/60, las fracciones concentrados del péptido de interés tras ser juntadas y concentradas en tampón fosfato sódico 50mM, pH 5.8 con NaCl 100 mM.

El test de Bradford permitió determinar un rendimiento de 20 mg de péptido purificado por un litro de cultivo. Se dejó el péptido de fusión con los residuos de histidina, ya que se consideró que esta cola no debería suponer un elemento que impidiese la actividad o interacción de la proteína. Por tanto, el sistema de expresión es ideal para la producción de este dominio ya que permite obtener significantes cantidades de proteína para llevar a cabo estos estudios.

2.2.2.2 Construcciones derivadas de la Lyk3-LysM-2 (de 1 a 7)

De manera similar a lo realizado con LysM-CAMs, las construcciones conteniendo el dominio Lyk3-LysM-2 se clonaron en plásmidos del tipo pRSET A y se usaron para transformar células C41. Se realizaron experimentos de expresión para todas las construcciones de Lyk3-LysM-2-(1-7). Después de la lisis bacteriana y la separación de las fracciones solubles y insolubles por centrifugación del lisado a 4000 rpm a 4°C durante 15 min, y se analizaron en SDS-PAGE al 15%.

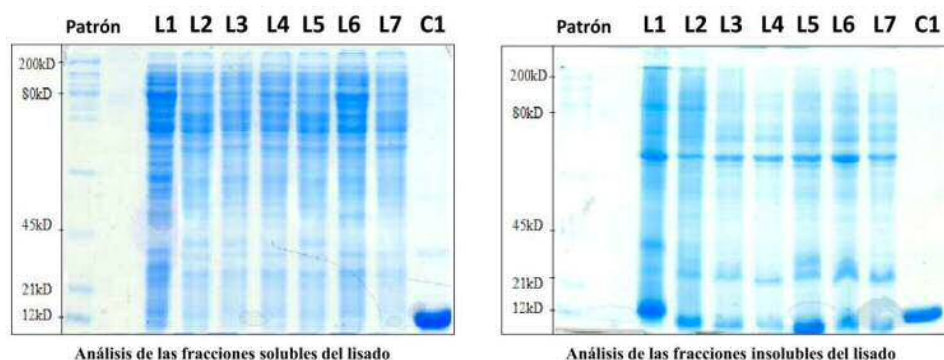


Figura 3.18 Geles de SDS-PAGE al 15% de alícuotas de las fracciones solubles (izquierda) e insolubles (derecha) de ensayos de expresión de las proteínas de fusión recombinantes del Lyk3-LysM2-(1-7) en minicultivos. (L(n) = Lyk3-LysM-(n) y C = LysM-CAM-1).

Los péptidos recombinantes se detectaron exclusivamente insolubles formando parte de los cuerpos de inclusión.

Puesto que no se pudieron obtener los péptidos recombinantes en forma soluble, se probaron otras condiciones empleando otros vectores y cepas bacterianas ya que la formación de cuerpos de inclusión puede ser influenciada por la naturaleza de las proteínas, la célula huésped, el nivel de expresión resultante del vector de elección y también las condiciones de inducción y crecimiento [88]. Cuando no se consiguió la mejora, se procedió a la solubilización de los agregados y su purificación a partir de los cuerpos de inclusión mediante el uso de tampones de solubilización basados en la combinación de detergentes y agentes reductores en condiciones desnaturizantes, utilizando altas concentraciones de urea o cloruro de guanidinio, fue posible solubilizar los péptidos obtenidos a partir de las construcciones LysM-2-Lyk-3-1 y LysM-2-Lyk-3-2. Todos los intentos de renaturalización de estos péptidos resultaron infructuosos.

2.2.3 Caracterización estructural del dominio LysM-CAM-1

A partir del péptido recombinante LysM-CAM-1 purificado, manteniendo el segmento adicional de 17 aa, se realizaron los experimentos 2D homonucleares de RMN con el objetivo de confirmar la estructura y el plegamiento previamente descritos por el dominio LysM-2 original del MltD de *E.coli* [55] tomando como referencia la asignación de los espectros descritos para el péptido natural y depositados en el BMRB (BioMagResBank: <http://www.bmrb.wisc.edu/>), código de acceso: BMRB4680.

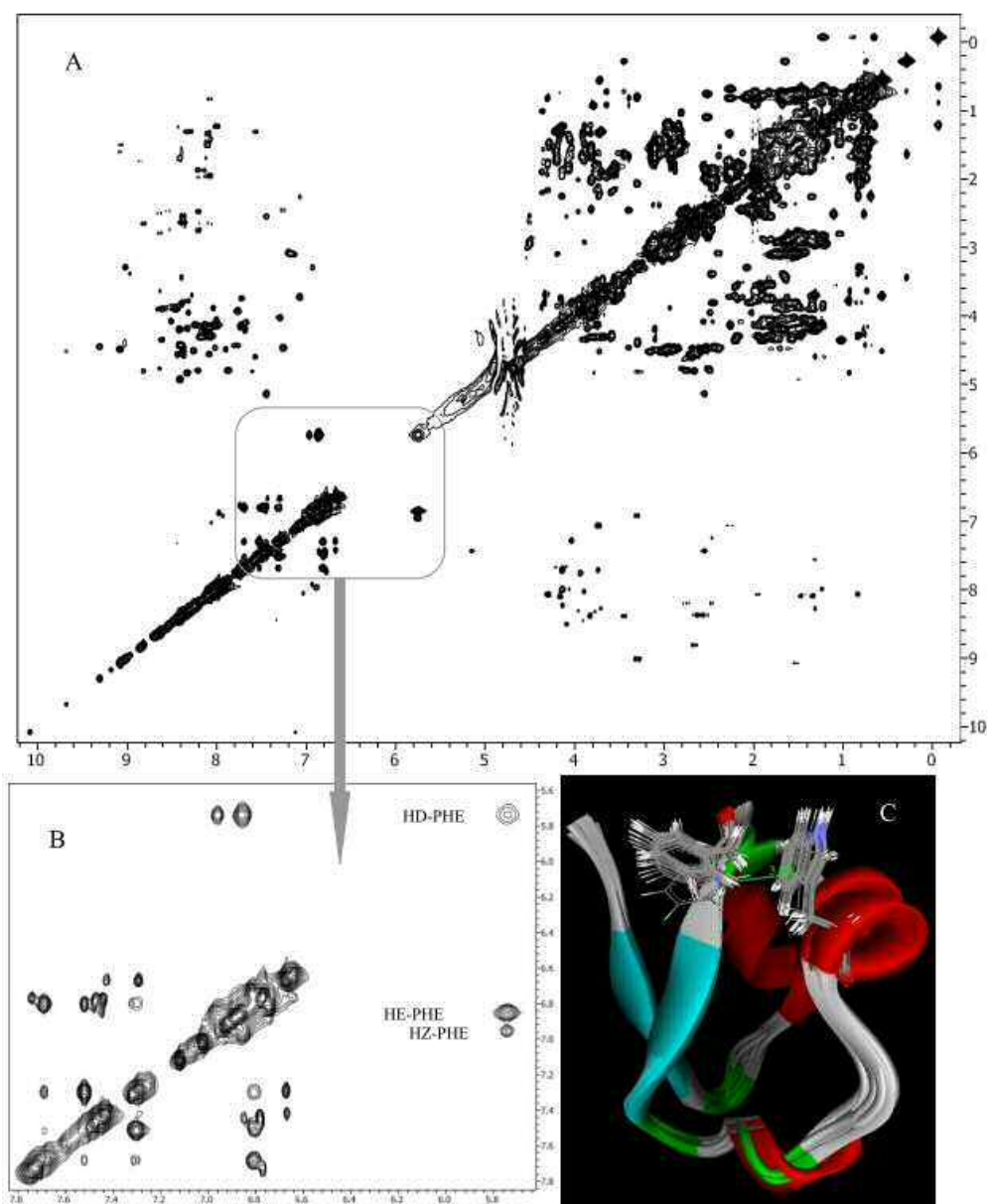


Figura 3.19 A) Espectro TOCSY (40 ms tiempo de mezcla) de CAM1. B) ampliación de la zona correspondiente a los protones aromáticos. C) Imagen de las 20 estructuras del dominio LysM de la transglucosidasa lítica (PDB 1E0G, [55]). Se resalta la cercanía entre la PHE 45 y el TRP 30 (numeración del dominio LysM, 48 aa totales).

En el espectro TOCSY, se indican los picos de correlación de las cadenas laterales de los aminoácidos aromáticos (Figura 3.19-B). Tras una asignación parcial incluyendo los picos de cruce registrados en la zona aromática, es interesante resaltar que el

desplazamiento químico de los protones PHE-D del anillo de la PHE-45 (Figura 3.20-A y B), que tienen un valor anómalo (más de 1 ppm desplazado hacia zonas de campo alto), se corresponden con los descritos para el dominio LysM de la proteína MltD original de bacteria (BMRB: 4680).

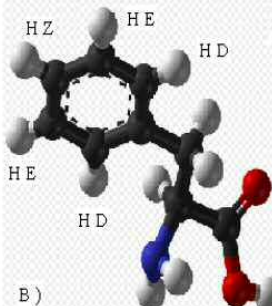
A)	Desplazamiento químico en (ppm)			
	HD	HE	HZ	
PHE (valor promedio)^a	7,22	7,28	7,28	B)
PHE BMRB4680^b	5,74	6,83	6,95	
PHE LysM-CAM-1	5,74	6,86	6,96	

Figura 3.20 A) Comparación de valores de desplazamiento químico de los HD HE y HZ de la fenilalanina -a) Datos tomados del BioMagResBank, NMR statistics (BMRB: <http://www.bmr.bu.edu/>). b) Datos para el dominio LysM de MltD [55] & B) Estructura de la fenilalanina.

Este desplazamiento químico anómalo es un buen diagnóstico del correcto plegamiento del péptido ya que se explica por el efecto apantallante que producen las corrientes de anillo del único triptófano presente en el dominio LysM. En la estructura depositada en el PDB (1E0G) para este dominio, se observa que ambas cadenas laterales están lo suficientemente cerca para justificar los desplazamientos químicos anómalos observados (Figura 3.19-C).

2.2.4 Estudios de interacciones entre LysM-CAM1 y carbohidratos

2.2.4.1 Interacción con quitina y derivados

La interacción entre el dominio LysM-CAM-1 y la quitina se evaluó utilizando tanto quitina soluble, pentaacetil-quitopentaosa (pentasacárido derivado de la quitina) como insoluble empleando una resina de quitina.

- Interacción del dominio LysM-CAM-1 con pentaacetil-quitopentaosa

Se siguió un protocolo de valoración por RMN de una muestra de LysM-CAM-1 a concentración constante (0,4 mM), en tampón de fosfato sódico pH 5,8 con 10% de D₂O, sobre la que se añaden cantidades crecientes de pentaacetil-quitopentaosa según se describe en el apartado de materiales y métodos, y se registraron una serie de 10 espectros de protón monodimensionales. En la figura 3.21 se presentan 4 puntos de la valoración.

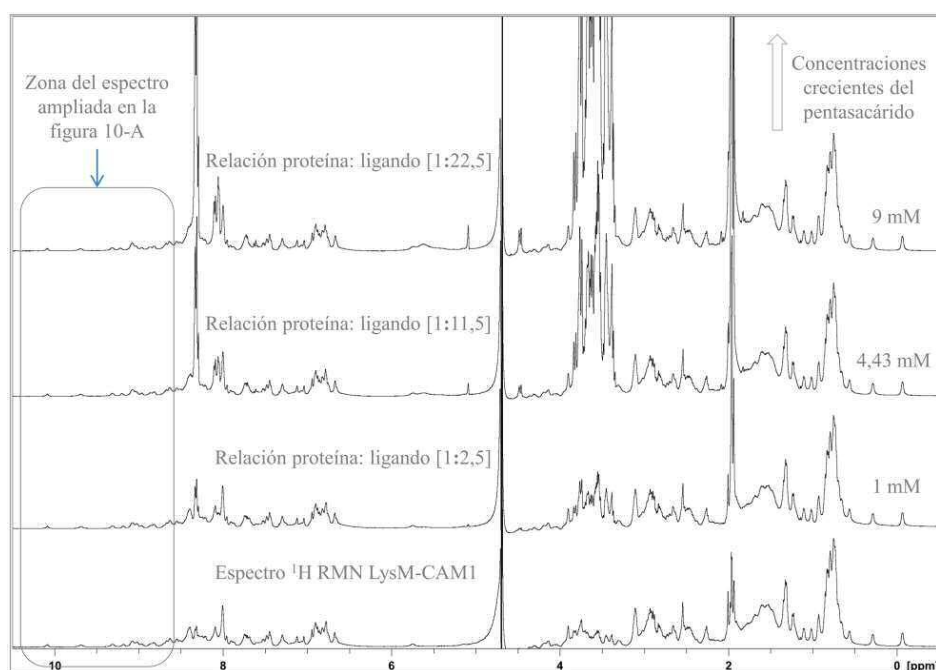


Figura 3.21 Serie de espectros de ¹H RMN (500MHz, pH 5.8, 298K) de la valoración a diferentes concentraciones (mM) del pentaacetil-quitopentaosa, con el péptido LysM- CAM-1 a 0.4 mM.

Los espectros se analizaron para detectar variaciones de desplazamientos químicos en alguna de las señales del péptido. Únicamente se observaron variaciones poco significativas (ninguna superior a 17 Hz con la máxima concentración de ligando, (9 mM, relación proteína/ligando 1:22,5) en unas pocas señales. En la figura 3.22-A se amplía la zona entre 8,8 y 10,2 ppm donde se observa el desplazamiento de algunas señales del péptido.

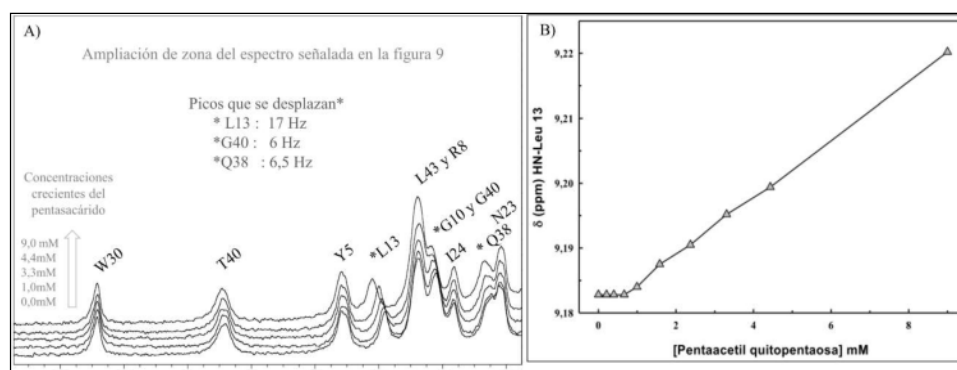


Figura 3.22 A) Serie de espectros de ^1H -RMN (500MHz, pH 5.8, 298K) de la valoración a diferentes concentraciones (mM) del pentasacárido, con el péptido LysM- CAM-1 a 0.4mM. Se identifican los H-N del esqueleto peptídico para diversos aminoácidos y el H-N de la cadena lateral del W30 **B)** variación de desplazamiento químico en función de la concentración del pentaacetil-quitopentaosa añadido.

Cuando se representó la variación de desplazamiento químico en función de la concentración del pentaacetil-quitopentaosa añadido (Figura 3.22-B) se observó una progresión lineal, en el rango de concentraciones usado. Esto indica que, de existir interacción, ésta debe ser de una afinidad muy baja e inespecífica ($K_d > 5 \text{ mM}$).

- Interacción del dominio LysM- CAM-1 con resina de quitina

Como la quitina natural es una biomacromolécula insoluble y heterogénea, se estudió la posible interacción del dominio LysM-CAM-1 con una resina comercial compuesta por partículas de quitina fácilmente manipulables en su forma hidratada. A una disolución de LysM-CAM-1 (0.4mM), en tampón fosfato sódico a pH 5,8, se añadieron alícuotas sucesivos de una suspensión de la resina de quitina (2mg/ml). El posible decaimiento de la intensidad de las señales por adhesión del LysM-CAM-1 a las partículas de quitina se siguió registrando sucesivos espectros de protón monodimensionales de la fracción soluble una vez sedimentada la resina (Figura 3.23).

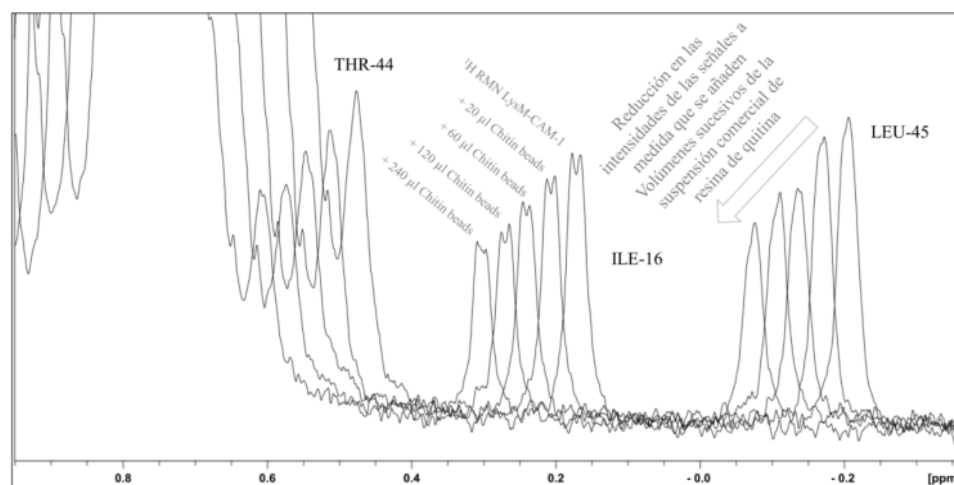


Figura 3.23 Serie de espectros de ^1H RMN (500MHz, pH 5.8, 298K) (lateralmente desplazados) de la valoración del espectro de LysM-CAM-1 a 0,4 mM añadiendo volúmenes sucesivos de la resina de quitina. Zona de campo alto donde se identifican las señales correspondientes a los metilos de algunos aminoácidos.

A medida que se aumenta el volumen usado de la resina de quitina, se observa un decaimiento en las intensidades de las señales del péptido LysM-CAM-1, que reduce la intensidad de las señales hasta 2/3. Esta reducción, se explica por la dilución de la muestra como se muestra en el siguiente gráfico (Figura 3.24). En este caso, el dominio LysM-CAM-1 no parece interactuar de manera eficiente con la quitina permaneciendo en solución sin ser extraído por las partículas de quitina.

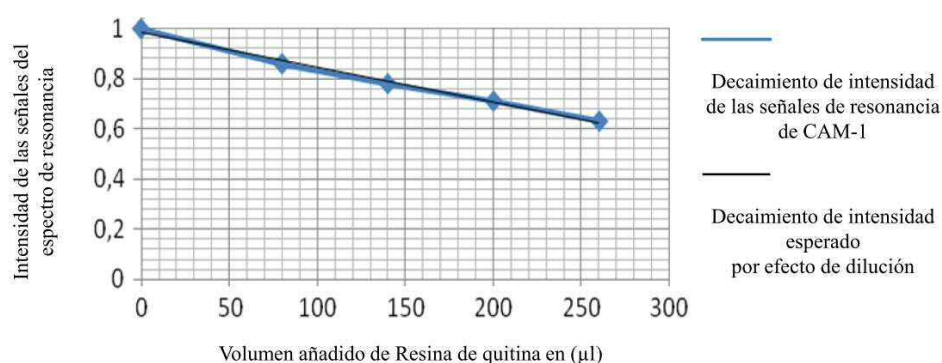


Figura 3.24 Representación de la disminución de la intensidad relativa de los señales del péptido LysM-CAM-1 en función de la cantidad de resina de quitina añadida al tubo de resonancia.

2.2.4.2 Interacción con peptidoglicano y derivados

Teniendo en cuenta que la proteína lítica MltD de *E. coli*, de la cual se ha tomado la secuencia del dominio LysM-CAM-1, se ha descrito con capacidad de unión a peptidoglicano [53-54, 89], también se estudió la posible interacción de este dominio con el propio peptidoglicano y con un fragmento derivado de éste, el GMDP (Figura 3.6) como modelo sencillo del peptidoglicano, que incorpora su unidad mínima disacáridica GlcNAc-MurNAc.

- Interacción del dominio LysM- CAM-1 con GMDP

Se realizaron una serie de adiciones sucesivas de alícuotas de una disolución stock del GMDP (a 11mM) sobre una muestra de LysM-CAM-1 a 0.4 mM, pH 5,8 a 25°C y hasta llegar a una relación 5:1 GMDP:LysM y se registraron los correspondientes espectros monodimensionales de protón.

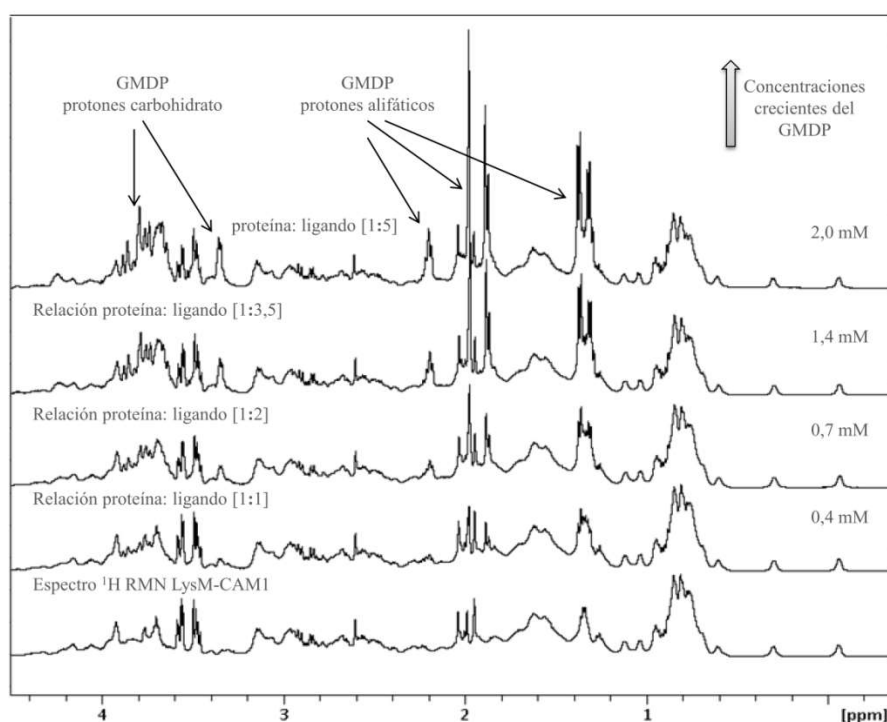


Figura 3.25 Serie de espectros de ^1H RMN (500MHz, pH 5.8, 298K) de la valoración a diferentes concentraciones (entre 0,4 y 2 mM) del GMDP, con el péptido LysM- CAM-1 a 0.4mM.

En estas condiciones experimentales no se observó ninguna perturbación significativa en las señales del espectro de la proteína que permitiese demostrar la interacción buscada (Figura 3.25).

- Interacción del dominio LysM-CAM-1 con el peptidoglicano

Alternativamente, se llevaron estudios de titulación por RMN para estudiar la interacción del propio peptidoglicano completo de *E. coli* con el dominio LysM-CAM-1. Aprovechando el carácter de polímero insoluble del peptidoglicano se siguió un protocolo similar al utilizado para estudiar la interacción del péptido con la resina de quitina. En la figura 3.26, se presenta la zona alifática donde aparecen señales aisladas para los metilos de la isoleucina 16 y leucina 45 cuyo significativo decaimiento no parece correlacionar con la dilución de la muestra con los sucesivos volúmenes añadidos de peptidoglicano.

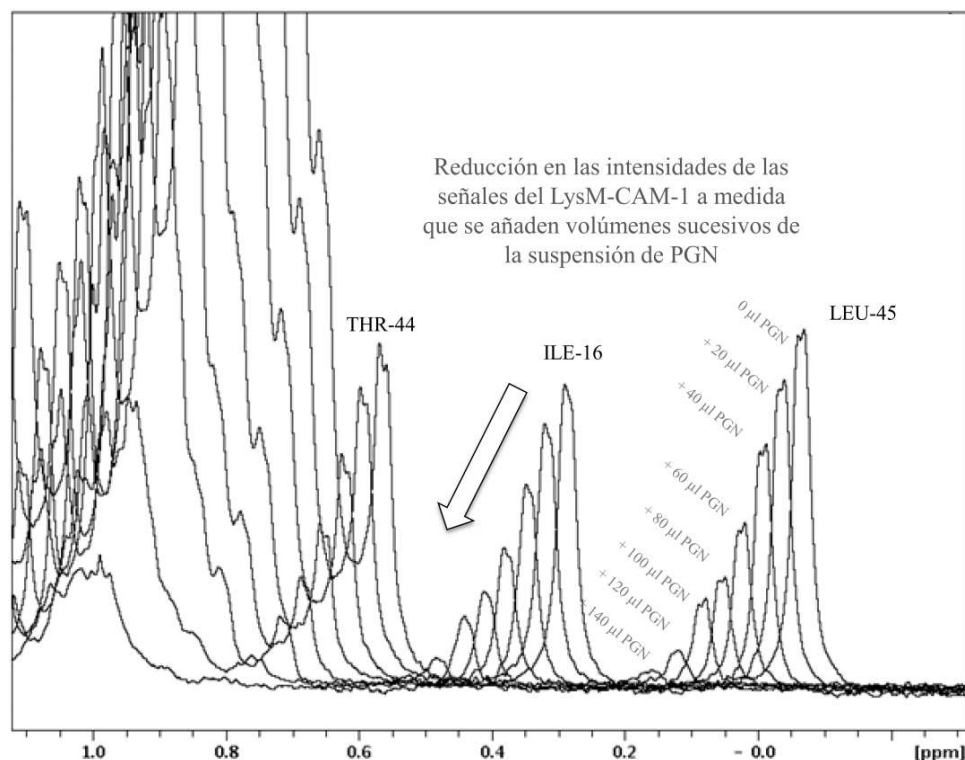


Figura 3.26 Serie de espectros de ^1H RMN (500MHz, pH 5.8, 298K) lateralmente desplazados de la valoración a diferentes concentraciones (mM) del peptidoglicano con el péptido LysM-CAM-1 a 0.4 mM. Zona de campo alto donde se identifican las señales correspondientes a los metilos de algunos aminoácidos.

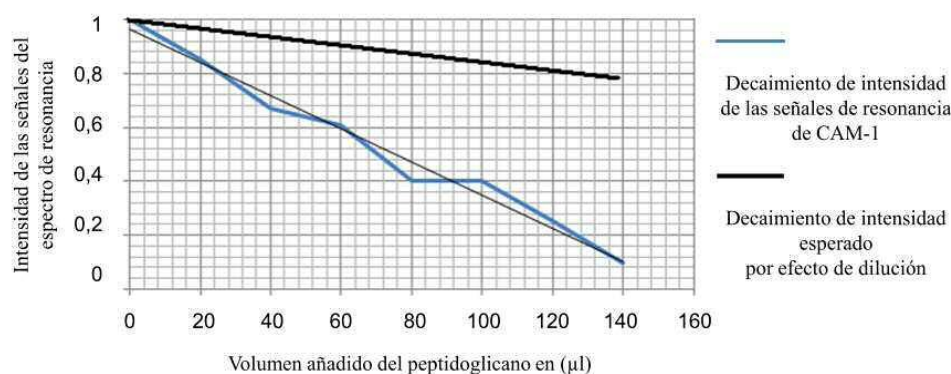


Figura 3.27 Representación gráfica de la caída del señal de Lys-M-CAM-1 debido a la dilución (línea de color negro) y debido a la extracción de proteína de la fase soluble por interacción con el peptidoglicano insoluble (color azul).

A partir de la representación gráfica del decaimiento, se deduce que la desaparición progresiva de la señal del péptido LysM-CAM-1 en función del aumento de la concentración del peptidoglicano no se puede justificar únicamente por la dilución de la muestra, como se observó en el caso de la resina de quitina (Figura 3.27).

El péptido LysM CAM-1 debe formar un complejo con el peptidoglicano que le hace inobservable por RMN a la hora de unión por sus características hidrodinámicas. El efecto neto es una extracción del péptido LysM-CAM-1 de la fase soluble por parte del peptidoglicano insoluble.

La implicación de los dominios LysM en la ruptura hidrolítica de las cadenas de GlcNAc o en el reconocimiento directo de residuos oligosacáridicos derivados de GlcNAc como los factores de nodulación y la participación en los sistemas de inmunidad y/o infección de plantas sigue siendo un campo muy activo de investigación [7, 9, 49-52, 81, 90]. Queda por definir, a escala atómico-molecular, si existen patrones generales de interacción entre estos dominios y sus ligandos oligosacáridicos. Estudios recientes han identificado en plantas un receptor del tipo, LysM RLK1, que juega un papel esencial en las respuestas de defensa frente a hongos en *Arabidopsis thaliana* [49, 61]. Aunque estos estudios revelaron que ciertas mutaciones en LysM RLK1 suprimían las respuestas de defensa contra los patógenos, no se llegó a describir, a nivel molecular, el modo de la percepción de la quitina por LysM RLK1, aunque se ha propuesto que es la base de estos mecanismos de defensa.

En esta tesis, se ha estudiado si un dominio LysM aislado, LysM-CAM1 derivado de MltD de *E. coli*, era capaz de unirse específicamente y directamente a la quitina, mediante el uso de resina de quitina y del oligosacárido derivado, pentaacetil quitopentaosa.

Los estudios de unión realizados entre el LysM CAM1 y la resina de quitina expresados en termino de péptido extraído de la fase soluble, se siguieron mediante la caída de la intensidad de sus señales correspondientes que pudieran reflejar la cantidad de péptido unido a la quitina. Sin embargo, la observación directa de los correspondientes espectros ^1H -RMN, solo reveló una caída moderada de las intensidades de los picos registrados, justificable por la simple dilución de la concentración del péptido en el tubo de resonancia (Figura 3.24). Por otra parte, la interacción del péptido con el pentaacetylquitopentaosa permitió observar variaciones moderadas en el desplazamiento químico de algunas señales del péptido, pero el comportamiento lineal respecto a la concentración añadida de oligosacárido observado en estas variaciones (Figura 3.22 A y B) indicaba la ausencia de una unión específica.

Cabe pensar que, en vez de un único dominio aislado LysM, la presencia del segmento extracelular completo de la proteína, o al menos del segmento conteniendo los múltiples dominios LysM, junto con la capacidad de los propios dominios LysM para formar complejos heteroméricos sean determinantes en la posible unión a fragmentos de quitina, como se ha descrito, por ejemplo, en el caso de la estructura cristalina de la proteína Ecp6 del hongo *Claustroduim* (código pdb: 4B8V), capaz de interaccionar con la quitina a través de la unión colectiva de sus dos dominios (LysM1 y LysM3) formando complejos de gran afinidad [70].

En la figura 3.28-B se comparan las secuencias del dominio LysM de la proteína MltD, presente en LysM-CAM 1, y las correspondientes a los dominios LysM 1 y LysM 3 de la proteína Ecp6 y se representa el modo de unión a un fragmento de quitina por estos dos últimos dominios, que entre ambos, encierran literalmente los fragmentos de la quitina en una estructura tipo sándwich (Figura 3.28-A, C y D)[70].

Además del complejo de Ecp6 con quitotetraosa, recientemente se han descrito la estructura de otros dominios LysM formando complejos con quitoooligosacáridos como el receptor de quitina AtERK1 (*Elicitor Receptor Kinase*, código pdb 4EBZ) de *Arabidopsis thaliana* [19] (Figura 3.14), una proteína de plantas implicada en la defensa

frente a hongos al reconocer la quitina. El modo de unión del ligando es similar en ambas proteínas (Figura 3.14, panel inferior) y se ha propuesto que también induce dimerización del receptor.

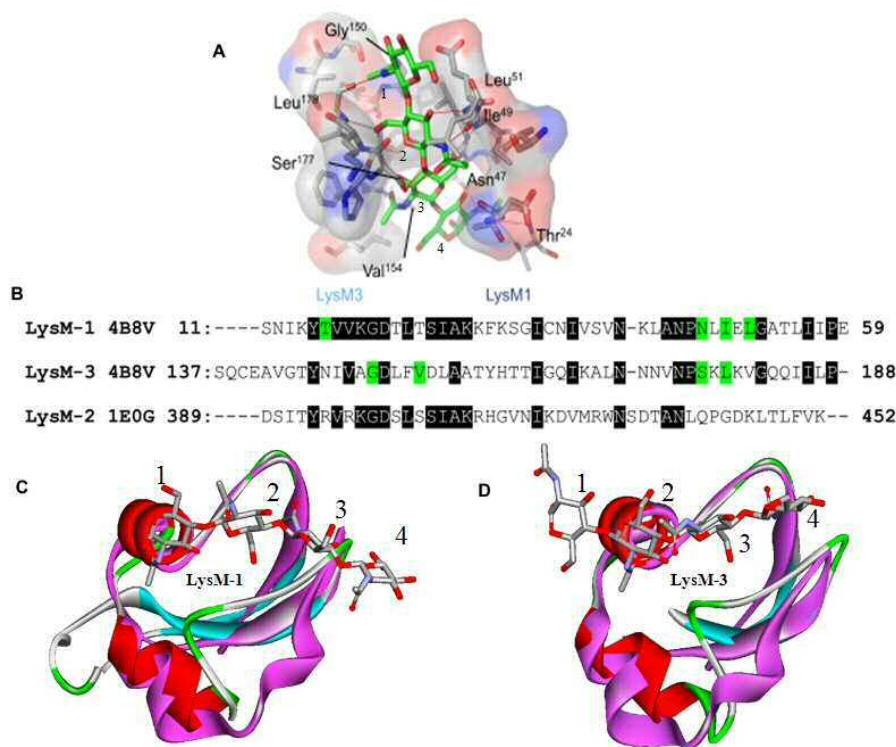


Figura 3.28 **A)** Representación del modo de unión del tetrasacárido derivado de quitina con los dominios LysM 1 y LysM 3 de la proteína Ecp6 (pdb: 4B8V). Se indican los principales residuos que forman enlaces de Hidrogeno con la quitina. **B)** Alineación de las secuencias de los dos dominios de LysM 1 y 3 de Ecp6 con el dominio LysM-2 de MltD, se muestra las similitudes entre las secuencias. En verde se resaltan los aa que participan en el sitio de unión. **C-D)** Superposicion de la estructura del segundo dominio LysM de MltD sobre la estructura cristalina de cada uno de los dominios LysM 1(C) y LysM 3 (D) de Ecp6 en color violeta y en complejo con el mismo tetrasacarido derivado de quitina.

A la vista de estos complejos quitooligosacárido-LysM, es interesante resaltar que la orientación en la que queda el tetrasacárido sobre el dominio LysM-2 de MltD, al superponer éste sobre las estructuras de los dominios LysM1 y LysM3 de Ecp6 (Figura 3.28-C), es distinta de la propuesta por Imberty y col. [81] obtenida por "docking" y modelado molecular. En cualquier caso, sí se puede observar una zona común de ocupación donde se sitúan los residuos del extremo reductor del oligosacárido (Figura 3.29).

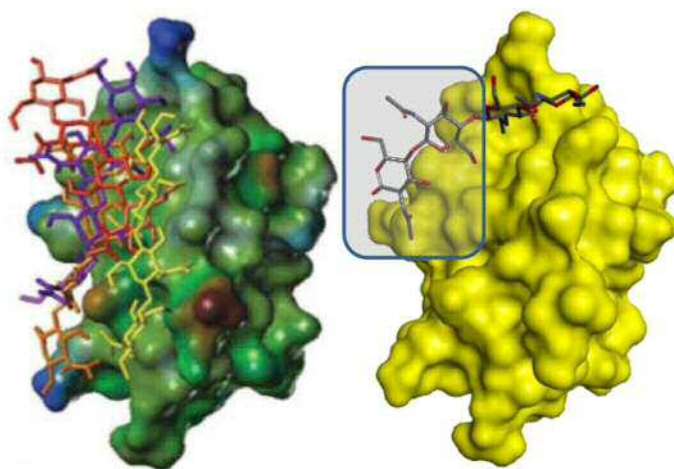


Figura 3.29 Comparación entre el modo de unión de la quitotetraosa propuesto sobre el dominio LyM-2 de MltD por cálculos de docking (izquierda) [81] y el modo de unión en la superficie del dominio LysM-1 de Ecp6 (Derecha), donde se resalta la zona común de ocupación donde se sitúan los residuos del extremo reductor del oligosacarido.

Por otro lado, como ya se ha mencionado, aunque la composición de la cadena polisacáridica del Peptidoglicano, del GMDP y de la quitina son similares, sólo se pudo detectar algún tipo de unión de LysM CAM-1 con el PGN completo, al menos en nuestras condiciones experimentales (Figuras. 3.26 y 3.27) y no con el fragmento GMDP. Este resultado muestra que la LysM-CAM-1, con un único dominio LysM, necesita cadenas oligosacáridicas con más de dos residuos. Además, no se excluye la participación de los aminoácidos del peptidoglicano en el proceso de unión para poder formar complejos con afinidades significativas. Este resultado no es incompatible con un estudio previo que sugirió que el sitio de reconocimiento de la quitina y del peptidoglicano pueden ser diferentes, como se ha descrito para los receptores presentes en la hoja de *Arabidopsis thaliana* [89]. De hecho, se conoce al menos cuatro dominios LysM RLKs [91]. Por lo tanto, puede especularse que algunos de ellos puedan estar involucrados en la unión al PGN y otros en la unión a fragmentos de quitina.

A la vista de los trabajos recientemente publicados, y de nuestros resultados experimentales, podemos concluir que la estrategia reduccionista inicialmente elegida para definir el sitio y modo de unión del ligando, no ha sido la más adecuada. Por el lado de la proteína. La utilización de un único dominio LysM puede ser insuficiente en

el caso de que la participación de dos (o más) dominios LysM sea una condición general para que se produzca una interacción eficiente en estos sistemas, e incluso para que los propios péptidos sean estructuralmente estables. Por el lado del ligando, también parece que compuestos sencillos de bajo peso molecular (GMDP) pueden no llegar a establecer suficientes contactos para formar complejos estables, indicando que los sitios de unión en la proteína deben ser extendidos.

En el momento de comenzar los trabajos incluidos en la presente Tesis se conocía, a nivel celular, la alta complejidad funcional de estos sistemas. Por ejemplo, diversos tipos de proteínas receptoras de plantas a las que se asociaban procesos de hetero- u homo-oligomerización de los receptores tanto para la detección de elicitores o patrones moleculares asociados a patógenos, PAMPs, como para la posterior señalización y activación/desactivación del sistema inmune en plantas. Sin embargo, no había información a escala atómico-molecular. En el momento actual, en el que ya se dispone de algunas estructuras 3D de complejos de estos dominios LysM con sus ligandos, se empieza a vislumbrar que la mencionada complejidad funcional va acompañada también de un nivel superior de complejidad en las interacciones intermoleculares subyacentes a estos procesos. De hecho, para que estos procesos se lleven a cabo, es posible que se requieran oligomerizaciones mediadas por el ligando. La comprensión a escala atómica de estos procesos de reconocimiento molecular sigue siendo un reto, tanto por las propias implicaciones funcionales de estos sistemas, como por las estrategias necesarias para su estudio.

2.3 CONCLUSIONES

Se ha confirmado experimentalmente por RMN que el dominio LysM derivado de la transglucosilasa lítica de *E. coli* (MltD) interacciona con el peptidoglicano, usando como modelo el péptido recombinante LysM-CAM-1 que incluye la secuencia LysM 389-452 de la proteína MltD con dos mutaciones S388C y N453C. Sin embargo, que no se observaron interacciones con el mínimo elemento disacárido, el GMDP, que se repite en la estructura del peptidoglicano.

3 DOMINIO PASTA

3.1 INTRODUCCIÓN Y ANTECEDENTES

En este apartado se presenta el estudio de otro tipo de dominios, denominados dominios PASTA (*PBP And Serine/Threonine kinase Associated domains*) al que inicialmente se le asoció una actividad de unión a antibióticos beta lactámicos [8, 92-93]. Estos dominios se han encontrado formando parte de proteínas homologas a las PBPs (*Penicillin Binding Proteins*) implicadas en el metabolismo del peptidoglicano bacteriano o de proteínas del tipo Ser/Thr quinasas presentes en las membranas de organismos procariotas, y que se clasifican por su homología con las quinasas eucariotas. A través de sus dominios PASTA extracelulares, desempeñan un papel fundamental en el control de diferentes procesos celulares bacterianos, tales como la regulación de rutas metabólicas y transducción de señales [94-96], la formación de biopelículas [97], la biosíntesis y división de la pared celular [98], la respuesta al estrés [99]. También están presentes en quinasas involucradas en el proceso de la esporulación que permite a las bacterias sobrevivir frente a las diversas condiciones medioambientales desfavorables como la escasez de nutrientes [97, 100-102].

Con el avance de la secuenciación de los genomas en bacterias, se han identificado más de 11000 secuencias diferentes de dominios PASTA que forman parte de proteínas bacterianas con arquitecturas moleculares muy variadas que van desde casos que incluyen un único dominio, hasta arquitecturas que llegan a agrupar 7 dominios PASTA en tándem (<http://pfam.xfam.org/family/PF03793>, base de datos pFam, entrada PF03793). En el genoma de *B. subtilis* y de *S. aureus* hay varios genes que codifican para proteínas con tres dominios PASTA como la quinasa PrkC que participa en la salida bacteriana del período de la latencia (Figura 3.30) [103-104].

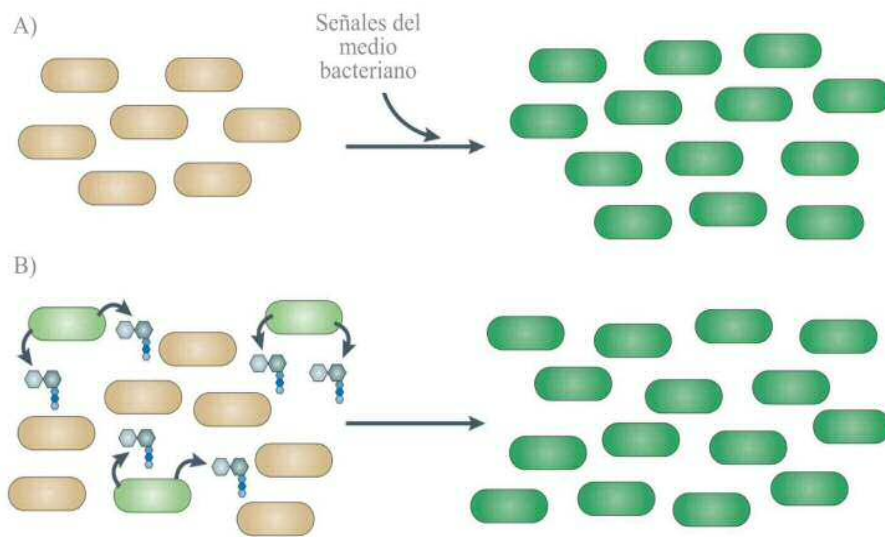


Figura 3.30 A) Esporas bacterianas (color marrón) en estado latente para sobrevivir en condiciones desfavorables y consisten en una notable reducción del metabolismo hasta el punto de reposo absoluto en algunos casos. La primera señal clave que desencadena la activación de la esporas (bacterias en color verde) es la presencia de nutrientes. **B)** Las células bacterianas también pueden salir de la latencia en respuesta a las señales liberadas por otras bacterias (Figura modificada de [105]).

En la fase de crecimiento de las bacterias se liberan fragmentos de PGN (muropéptidos) en el medio, debido a la actividad de distintas enzimas llamadas autolisinas (hidrolasas y amilasas) capaces de degradar y hidrolizar el propio peptidoglicano de la bacteria en ciertos enlaces en los que se incluyen nuevas moléculas que intervienen en el proceso de elongación de la pared [106-108]. Estos fragmentos es uno de los factores que indica la presencia de condiciones favorables en el medio para las bacterias inactivas y se asocia con el proceso de reactivación bacteriana (Figura 3.30-B).

Los fragmentos de PGN liberados (muropéptidos), dependiendo del aminoácido situado en la tercera posición de su cadena peptídica, pueden ser de tipo Lisina (Lys) o de tipo ácido meso-diaminopimérico (DAP) [105, 109] (Figura 3.31).

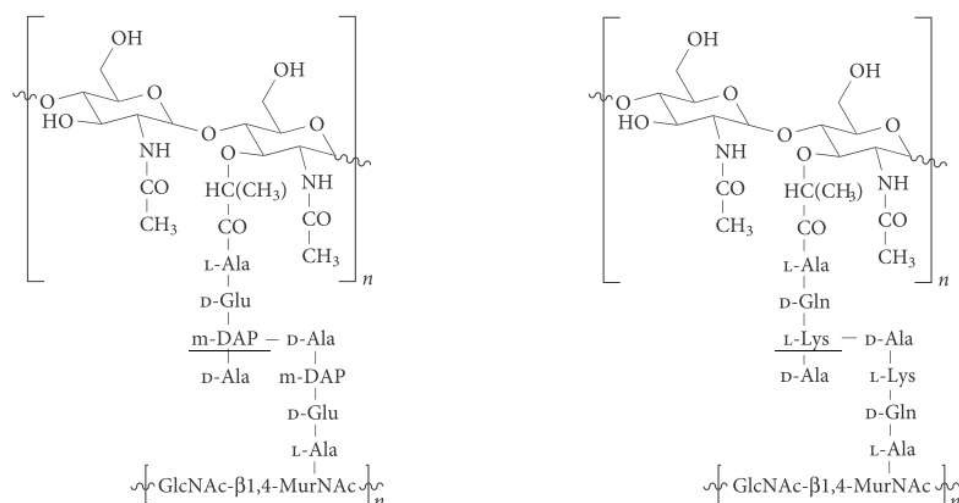


Figura 3.31 Los dos tipos de muropéptido que compone la pared celular bacteriana. La mayoría de las bacterias Gram positivas contienen un residuo L-Lys en la tercera posición del tetrapéptido (izquierda). Las bacterias Gram negativas, tienen siempre un residuo DAP en esta posición (derecha).

A nivel estructural, la organización de la proteína PrkC de *B. subtilis* es muy similar a la PrkC de *S. aureus* [110]. El dominio N-terminal de PrkC se encuentra en el citoplasma con actividad quinasa altamente conservado y continúa con una región transmembranal de 28 residuos y una porción C-terminal extracelular, con 3 dominios PASTA (Figura 3.32) [111-112].

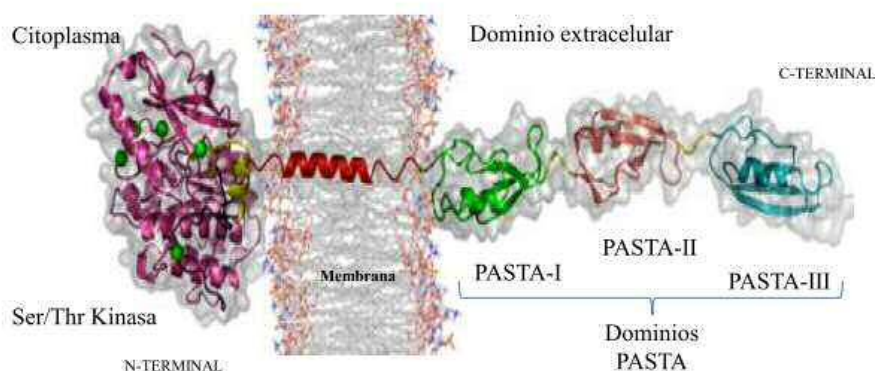


Figura 3.32 Representación de organización estructural de la proteína PrkC de la pared celular de *S. aureus*, homóloga a la de *B. Subtilis*. La organización comprende un dominio intracelular quinasa N-terminal y un dominio C-terminal extracelular, con 3 dominios PASTA consecutivos y un dominio de tipo inmunoglobulina (no representado en el extremo C-terminal), unidos por un segmento transmembranal de unos 28 aa. (Figura modificada de [111]).

Estudios previos de PrkC con el PGN indicaron que las esporas de *B. subtilis* germinan en respuesta a muropéptidos de tipo DAP, que constituyen la pared celular de las mismas bacterias, pero no frente a muropéptidos de tipo Lys. Esta observación sugirió que el dominio extracelular de la proteína PrkC de *B. subtilis* presenta una especificidad selectiva hacia muropéptidos de tipo DAP distinta a la asociada a la PrkC de *S. aureus* [109]. Estudios de interacción mediante STD revelaron que la unión de PrkC al muropéptido se produce principalmente a través de las cadenas laterales de la cola peptídica, en particular, a través del residuo de DAP (la máxima intensidad de STD registrada corresponde al protón H β -DAP). En cambio, los protones de la región de los carbohidratos interaccionan muy débilmente [112].

Por otra parte, las diversas estructuras tridimensionales de complejos entre otras proteínas con ligandos que contienen DAP, el grupo carboxílico del DAP forma un puente salino con la cadena lateral de un residuo de arginina (Figura 3.33). Este hecho sugiere que la unión del muropéptido se produce a través de la interacción un residuo de arginina del EC-PrkC con el DAP del muropéptido.

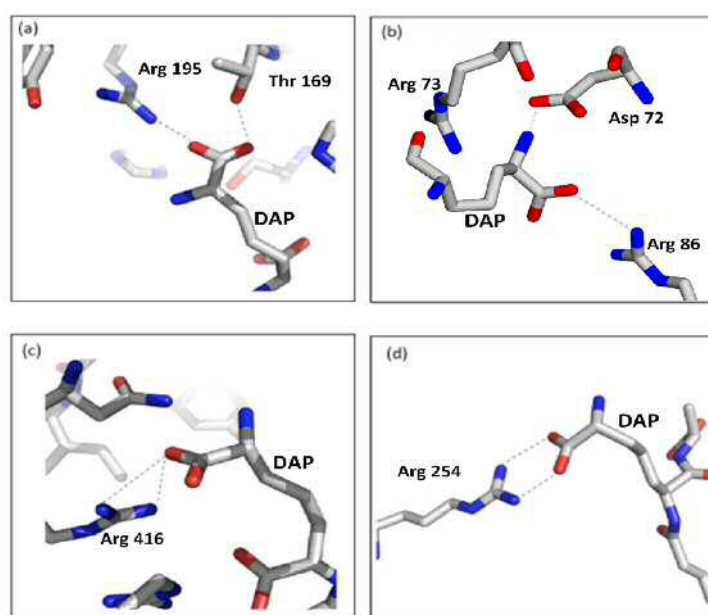


Figura 3.33 Interacción del residuo DAP en estructuras de complejos con proteínas disponibles en el Protein Data Bank.: **a)** diaminopimelato deshidrogenasa *Corynebacterium glutamicum* (código PDB 2DAP), **b)** lipoproteína de *Haemophilus influenza* (código PDB 2AIZ), **c)** UDP-N-acetylmuramoyl-L-alanil-D-glutamato: meso-ligasa de diaminopimelato *E. coli*, (código PDB 1E8C), **d)** Proteína de reconocimiento del peptidoglicano de *Drosophila* LE (PGRP-LE) (código PDB 2CB3)

En la secuencia de la parte extracelular de la PrkC de *B. subtilis*, sólo existen dos residuos de arginina: La Arg614, que está en el dominio tipo inmunoglobulina, que mediante modelos de homología, se ha propuesto que participa en un puente salino con Glu604. El otro residuo, la Arg500, queda expuesta al disolvente en el dominio PASTA3. Además, las mutaciones de Arg500 a alanina (R500A) o por ácido glutámico (R500E) impiden la unión a muropéptidos [110]. Estos datos llevan a pensar que la Arg500 pueda estar implicada en la unión al muropéptido de un modo similar a los representados en la figura 3.33.

10	20	30	40	50	60
MLIGKRISGR	YQILRVIGGG	GMANVYLAED	IILDREVAIK	ILRFDYANDN	EFIRRFRRREA
70	80	90	100	110	120
QSASSLDHPN	IVSIYDLGEE	DDIYYIVMEY	VEGMLTKEYI	TANGPLHPKE	ALNIMEQIVS
130	140	150	160	170	180
AIAHAHQNQI	VHRDIKPHNI	LIDHMGNIKV	TDFGIATALS	STTITHTSNV	LGSVHYLSPE
190	200	210	220	230	240
QARGGLATKK	SDIYALGIVL	FELLTGRIPF	DGESAVSIAL	KHLQAETPSA	KRWNPSVPQS
250	260	270	280	290	300
VENIILKATA	KDPFHRHYETA	EDMEADIKTA	FDADRLNEKR	FTIQEDEEMT	KAIPPIKDEE
310	320	330	340	350	360
LAKAAGEKEA	EVTTAQENKT	KKNGKRKKWP	WVLLTICLVF	ITAGILAVTV	FPSLFMPKDV
370	380	390	400	410	420
KIPDVSGMEY	EKAAGLLEKE	GLQVDSEVLE	ISDEKIEEGL	MVKTDPKADT	TVKEGATVTL
430	440	450	460	470	480
YKSTGKAKTE	IGDVTGQTVD	QAKKALKDQG	FNHVTVNEVN	DEKNAGTVID	QNPSAGTELV
490	500	510	520	530	540
PSQDQVTLTV	SI ^R GPEDITL ^R	DLKTYSKCAA	SGYLEDNGLK	LVEKEAYSDD	VPEGQVVKQK
550	560	570	580	590	600
PAAGTAVKPG	NEVEVTFSLG	PEKKPAKTVK	EKVKIPYEPE	NEGDELQVQI	AVDDADHSIS
610	620	630	640	648	
DTYEEFKIKE	PTERT ^R IELKI	EPGQKGYQV	MVNNKVVSYSK	TIEYPKDE	

Figura 3.34 Secuencia de la proteína PrkC de *Bacillus subtilis* (UniProtKB O34507). En verde se indica la región extracelular (351-648) que incluye los 3 dominios PASTA-1 (356-424, gris claro), PASTA-2 (425-492, gris oscuro) PASTA-3 (493-559, amarillo) y dominio tipo inmunoglobulina (629-648). Las dos argininas se muestran en color rojo.

En este trabajo se pretende demostrar, mediante RMN que, en la unión del PGN al tercer dominio PASTA de *B. subtilis*, el residuo R500 es un elemento crítico. Más allá, podría estar involucrado en la capacidad de la PrkC para discriminar entre los peptidoglicanos de tipo DAP y tipo Lys. Con este fin, se ha llevado adelante la determinación de la estructura tridimensional en solución, mediante RMN, del tercer dominio recombinante PASTA de *B. subtilis* para, a continuación, tratar de caracterizar los aminoácidos que participan en la interacción con diferentes ligandos potenciales, derivados del peptidoglicano y antibióticos betalactámicos.

3.2 RESULTADOS Y DISCUSIÓN

3.2.1 Expresión y purificación del dominio

El tercer dominio PASTA de la región extracelular de *B. subtilis* (residuos 493-559 de CE-PrkC, junto con 3 aminoácidos adicionales en el extremo C-terminal (GPE, hasta el 562), y cuatro aminoácidos adicionales (GAMG) en el extremo amino terminal, se produjo mediante expresión tanto en medio rico con abundancia isotópica natural como en medio mínimo enriquecido con isótopos estables N^{15} y C^{13} [109].

3.2.2 Determinación de la estructura 3D del dominio recombinante PASTA

En primer lugar, se usó la proteína no marcada isotópicamente para la adquisición de una serie de espectros homonucleares (TOCSY y NOESY) en diversos tiempos de mezcla, que permitieron identificar y asignar un número significativo de aminoácidos del péptido (Figura 3.35). Sin embargo, la presencia de cinco prolina y algunas pequeñas repeticiones internas en la secuencia implicando a varios restos de glutámico dificultaron la asignación completa de la secuencia (GPE 493-5, PE, 532-3 y GPE 560-2, ED 495-6 y 515-6, AA 509-510 y 542-3, KEA 507-9 y 524-6, VE 522-3 y EVEV 552-5).

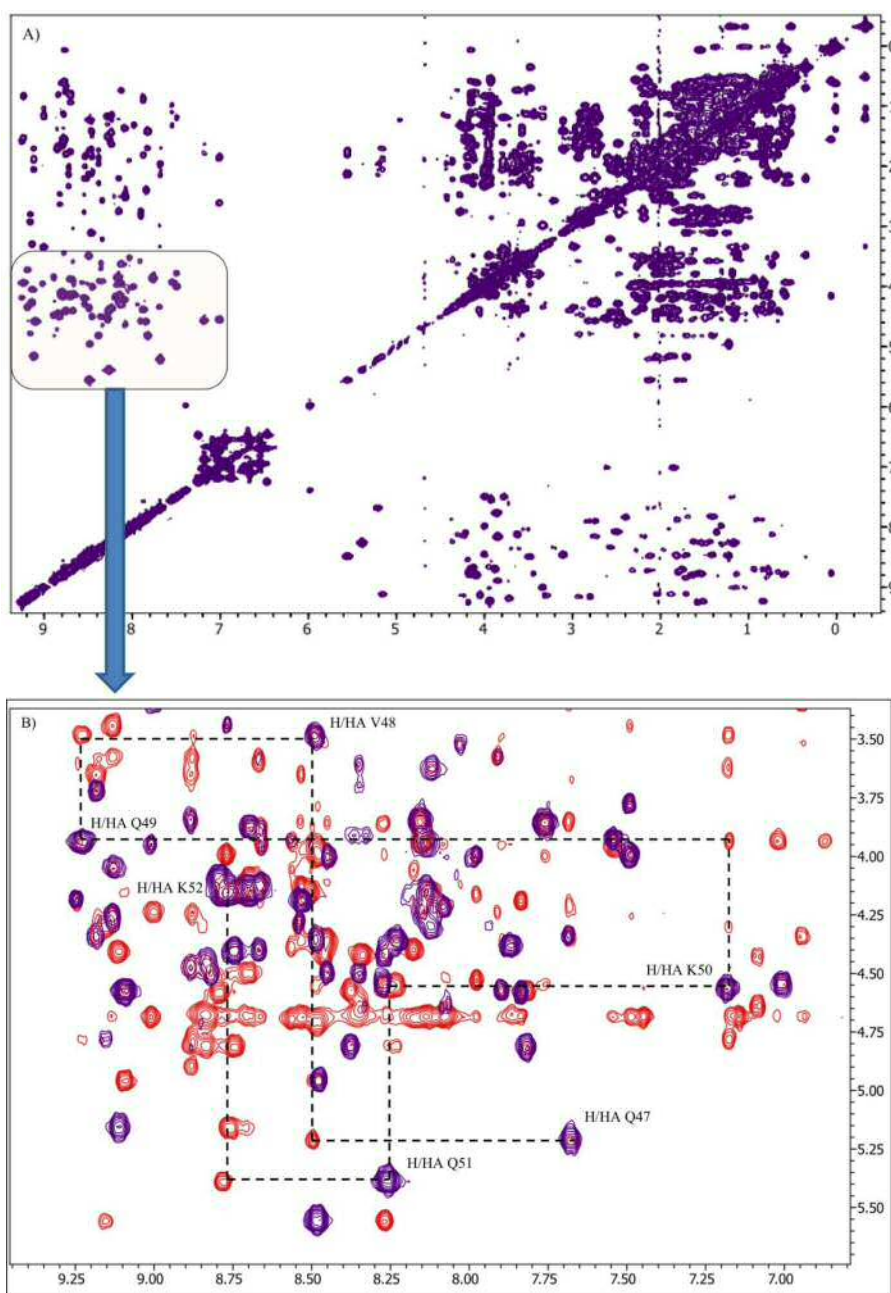


Figura 3.35 Espectros homonucleares empleados para la asignación secuencial del tercer dominio PASTA de PrkC de *B. subtilis*. **A)** Imagen del espectro ^1H - ^1H TOCSY. **B)** Superposición de los espectros TOCSY (color violeta) y NOESY (color rojo) que permitió identificar la conectividad entre los sistemas de espín comprendidas entre los residuos Q47-K52 (numeración de los aa (ver tabla 2)). Los espectros se adquirieron en un espectrómetro Bruker 600 MHz a 298K, empleando un tiempo de mezcla de 40 ms para ^1H - ^1H TOCSY y 200 ms para ^1H - ^1H NOESY. La proteína se utilizó a una concentración de 0.8 mM en tampón PBS, pH 7.3 con 10% D_2O .

La disponibilidad de la proteína marcada isotópicamente con ^{15}N , o con ^{15}N y ^{13}C , permitió aplicar estrategias de asignación mediante el uso de experimentos heteronucleares 2D y 3D del tipo: HSQC, HNCA, HN(CO)CA, TOCSY-HSQC, NOESY-HSQC, CBCA(CO)NH, HNCACB (Tabla 3.5). Algunos de estos estudios se realizaron en colaboración con la Dra. Marta Bruix del Instituto Rocasolano de Química-Física del CSIC en Madrid (España).

Tabla 3.5 Experimentos homo- y heteronucleares de RMN realizados sobre el tercer dominio PASTA de PrcK de *B. subtilis*

Experimento		Objetivo del experimento
Proteína sin marcar	2D TOCSY homonuclear	Permite identificar los sistemas de spin para cada aminoácido que conforman la molécula y asignar protones de cadena lateral, siempre y cuando el residuo no esté solapado con otros residuos.
	2D NOESY homonuclear	Permite la asignación secuencial de los aminoácidos de la cadena peptídica y es la base experimental para generar la colección de restricciones de distancia necesarias para el cálculo estructural.
Proteína isotópicamente marcada con ^{15}N	HSQC	Identificar los NH tanto de la cadena principal como la cadena lateral de los aminoácidos de la proteína
	TOCSY-HSQC	Asignar los protones de H^n de cada residuo.
	NOESY-HSQC	Identificar las restricciones de las distancias interprotones para los cálculos de la estructura, y asignamiento secuencial
Proteína doblemente marcada con ^{13}C y ^{15}N	HNCA	Asignar los $\text{C}\alpha$ a partir del ^1H - ^{15}N HSQC.
	HN(CO)CA	Asignar los $\text{C}\alpha$ y $\text{C}\alpha$ -1 a partir del ^1H - ^{15}N HSQC.
	HNCACB	Asignar los $\text{C}\alpha$ y $\text{C}\beta$ a partir de ^1H - ^{15}N HSQC
	CBCA(CO)NH	Asignar los $\text{C}\alpha$, $\text{C}\alpha$ -1 y $\text{C}\beta$ y $\text{C}\beta$ -1 a partir de ^1H - ^{15}N HSQC

Los espectros homonucleares de la proteína sin marcaje isotópico y los heteronucleares con ^{15}N se realizaron sobre muestras de dominio PASTA (a 0.8 mM en PBSx1 10% D_2O a pH:7.3) en espectrómetro Bruker 600 MHz a 298K, mientras que los experimentos heteronucleares del dominio PASTA doblemente marcado a 0.9 mM en tampón PBS con 10% D_2O , pH 6,5 se realizaron en un espectrómetro Bruker 800 MHz a 300K.

3.2.3 Asignación de las señales del esqueleto

La asignación específica de los δ de las parejas NH y de sus carbonos alfa ($C\alpha$) correspondientes al dominio PASTA se llevó a cabo mediante el uso combinado de los experimentos de triple resonancia HNCA y HN(CO)CA (Figura 3.36).

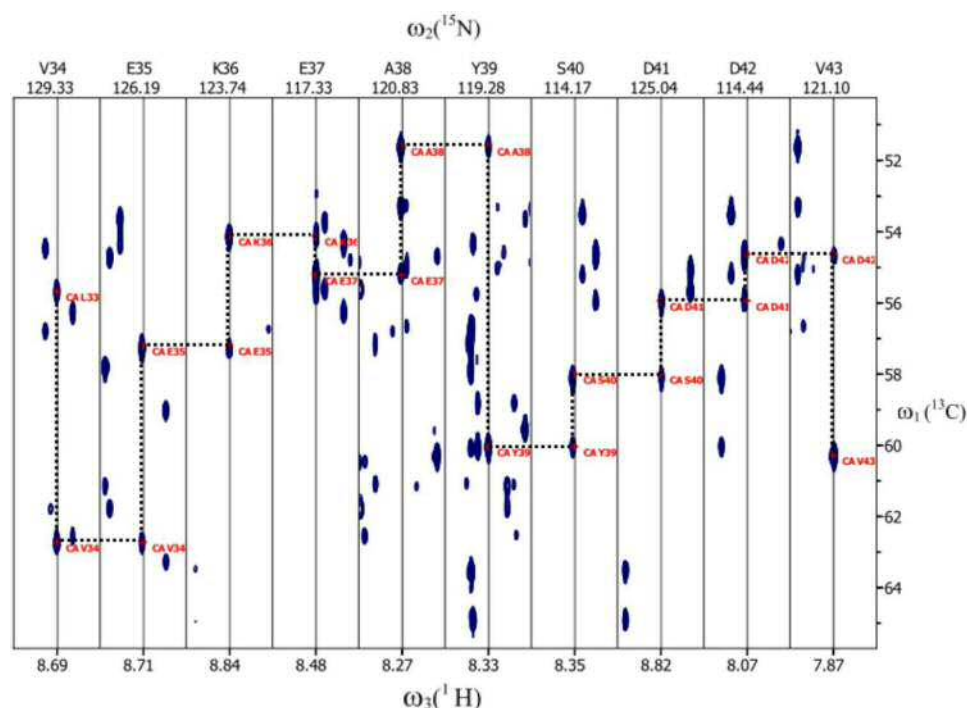


Figura 3.36 Colección de planos del espectro de 3D-HNCA correspondientes a las frecuencias de ^{15}N de los residuos V34-V43, que permite la asignación secuencial de los átomos de carbono α de los residuos mencionados mediante las conectividades $^1\text{H}_{\text{N}(i)}-^{15}\text{N}_{(i)}-^{13}\text{C}_{\alpha(i)}$ y $^1\text{H}_{\text{N}(i)}-^{15}\text{N}_{(i)}-^{13}\text{C}_{\alpha(i-1)}$ el espectro fue adquirido en un espectrómetro Bruker 800 MHz a 300K con una muestra triplemente marcada del dominio PASTA a 0.9 mM en tampón fosfato sódico 10%D₂O a pH 6,5.

En estos experimentos, la conexión secuencial entre los sistemas de espín se debe a los acoplamientos escalares que existen entre las parejas de núcleos $^1\text{H}-^{15}\text{N}$ y $^{15}\text{N}-^{13}\text{C}$. Estos acoplamientos, a uno o dos enlaces, no son dependientes de la conformación, con lo que el proceso de asignación es menos ambiguo que aquel que se realiza mediante los experimentos que emplean el efecto NOE. En este caso, la intensidad de los picos de cruce depende de la distancia entre los núcleos ($\text{NOE} \sim r^{-6}$). La distancia interprotónica, a su vez, depende enormemente de la conformación. Es decir, no siempre se observan todas las conexiones entre parejas de protones correspondientes a aminoácidos

secuenciales, o bien estas tienen menor intensidad que otras que corresponden a parejas de protones alejados en su estructura primaria. En el HNCA se observan las conectividades $^1\text{HN}_{(i)}\text{-}^{15}\text{N}_{(i)}\text{-}^{13}\text{C}_{\alpha(i)}$ y $^1\text{HN}_{(i)}\text{-}^{15}\text{N}_{(i)}\text{-}^{13}\text{C}_{\alpha(i-1)}$, mientras que en el HN(CO)CA solo se observan las conectividades $^1\text{HN}_{(i)}\text{-}^{15}\text{N}_{(i)}\text{-}^{13}\text{C}_{\alpha(i-1)}$. La combinación de ambos experimentos permite, por tanto la asignación de los δ de todos los espines involucrados.

Los δ de C_{α} están en el rango de 40 a 70 ppm y dependen del tipo de estructura secundaria presente en la proteína, en orden creciente de δ : lámina β < estructura no periódica < hélice α que además, los δ de C_{α} de los residuos de glicinas son los más bajos (40-47 ppm), mientras que los de prolina, los más altos (60-65 ppm) respectivamente.

Los valores de los δ de C_{α} del dominio PASTA-3 determinados se encuentran entre 43 y 69 ppm (Figura 3.37). Pudiendo distinguir 9 residuos de glicinas por sus δ de C_{α} característicos (43.60-47.71 ppm) y 5 prolinas con δ entre 62,77 y 64,52 ppm.

En el espectro $^{15}\text{N}\text{-}^1\text{H}$ -HSQC del dominio PASTA-3 se registraron 73 señales de correlación heteronuclear y una elevada dispersión desde 5.9 a 9.25 ppm, característica de la estructura secundaria en forma de lámina- β descrita para los dominios PASTA de *S. aureus* (Figura 3.37) [111]. La mayor parte de los picos de cruce NH esperados y que no aparecen en el HSQC fueron localizados más tarde como correspondientes a los N-terminal de la proteína. Los extremos de las cadenas polipeptídicas presentan normalmente una elevada movilidad y el acceso de moléculas de agua en el medio es relativamente fácil, por lo que existe un proceso de intercambio químico rápido entre los NH correspondientes y el agua de forma que los picos de cruce disminuyen en intensidad o incluso se pierden.

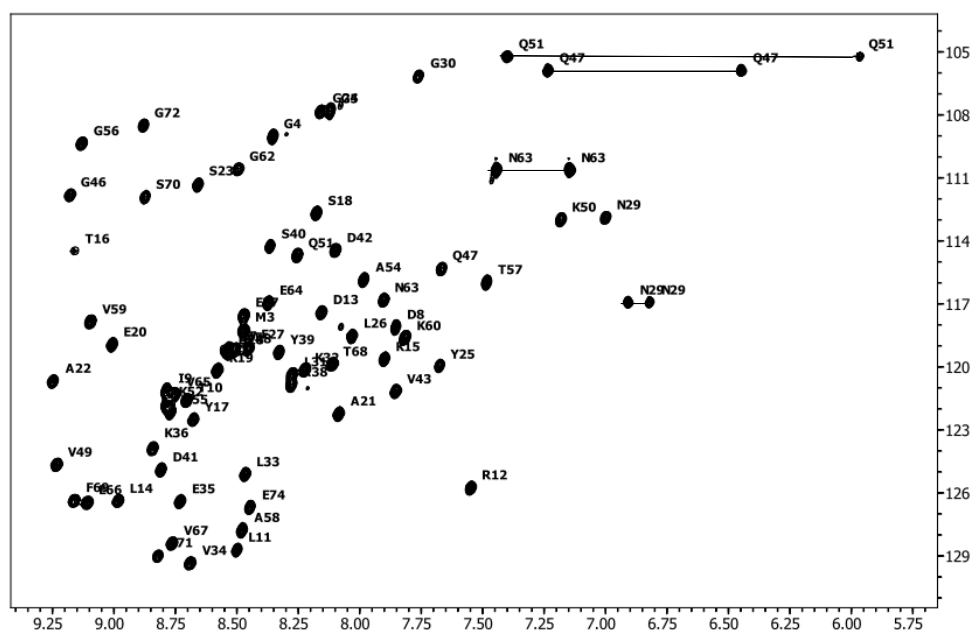


Figura 3.37 Espectro ^1H - ^{15}N -HSQC del dominio PASTA-3 recombinante de PrkC de *B. subtilis*. El espectro fue adquirido en un espectrómetro Bruker 600 MHz, a 298K, utilizando una concentración de péptido marcado con ^{15}N a 0.6 mM, en tampón PB 10% D_2O , pH 7.3. Los picos de cruce ^1H - ^{15}N amida se indican con el correspondiente número de residuo en la secuencia de la proteína.

El espectro HSQC también muestra las señales de los grupos $^{15}\text{N}^1\text{H}_2$ de las cadenas laterales de las dos asparraginas y las dos glutaminas del dominio recombinante. Estas señales son relativamente fáciles de identificar, ya que cada el ^{15}N correlaciona con dos ^1H unidos. Estos picos están indicados mediante líneas horizontales en la figura 3.37.

Una vez asignado el esqueleto del dominio, mediante experimentos 2D- TOCSY y 2D- NOESY y sus versiones 3D (3D-TOCSY-HSQC, 3D-NOESY-HSQC), se procedió a asignar los protones de las cadenas laterales. La tabla 2 resume los desplazamientos químicos del dominio PASTA-3.

Tabla 3.6 Desplazamientos químicos de las señales de ^1H ^{13}C y ^{15}N asignadas para el dominio PASTA-3 de PrcK de *B. subtilis*.

Nº de aa en la secuencia (fig. 5)	Residuo nº de aa en el péptido	N	H	CA	HA HA2/ HA3	CB	HB HB2/ HB3
--	GLY-1			43.6	3.76		
--	ALA-2			52.89	4.27	19.36	1.32
--	MET-3	118.22	8.47	55.7	4.43	32.87	2.05/1.95
--	GLY-4	109.01	8.36	45.34	3.92		
493	GLY-5	107.81	8.12	44.66	4.15/3.95		
494	PRO-6			63.62	4.13	32.38	2.29
495	GLU-7	119.12	8.53	56.55	4.19	30.63	2/1.84
496	ASP-8	118.09	7.83	54.85	4.58	41.82	2.39
497	ILE-9	121.13	8.78	60.36	4.1	40.51	1.65
498	THR-10	121.58	8.7	62.49	4.15	68.46	3.88
499	LEU-11	128.7	8.5	56.27	3.96	43.16	1.6/1.43
500	ARG-12	125.66	7.54	56.25	3.93	32.3	0.98/1.11
501	ASP-13	117.31	8.14	54.23	4.23	40.5	2.75/2.2
502	LEU-14	126.35	9	53.62	3.34	42.33	1.17/1.03
503	LYS-15	119.55	7.91	59.56	3.57	32.15	1.67/1.84
504	THR-16	114.1	9.13	65.73	4.04	69.87	4.48
505	TYR-17	122.5	8.67	58.81	4.4	38.13	3.59/3.03
506	SER-18	112.67	8.18	56.97	4.67	65.62	4.06/4.39
507	LYS-19	120.11	8.56	61.03	3.94	31.74	1.46/1.25
508	GLU-20	118.88	9.01	60.21	3.95	29.4	1.99/1.87
509	ALA-21	122.22	8.08	55.03	4.22	19.27	1.36
510	ALA-22	120.66	9.24	55.86	4.18	18.37	1.58
511	SER-23	111.25	8.66	62.6	4.12	63.24	3.96/3.89
512	GLY-24	107.79	8.16	47.71	3.85		
513	TYR-25	119.99	7.68	61.09	4.34	38.07	3.33/2.98
514	LEU-26	118.51	8.03	58.73	3.51	40.06	1.9/1.14
515	GLU-27	119.04	8.45	59.95	4	29.51	2.13/2.05
516	ASP-28	119.3	8.54	57.2	4.28	40.51	2.76/2.48
517	ASN-29	112.87	7	53.82	4.55	41.08	2.61/1.85
518	GLY-30	106.13	7.76	47.08	3.86		
519	LEU-31	120.35	8.27	53.15	4.55	43.76	1.5/1.16
520	LYS-32	120.15	8.23	55.05	4.36	34.25	1.44
521	LEU-33	125.03	8.45	55.67	4.51	42.6	1.85/1.5
522	VAL-34	129.33	8.69	62.72	4.15	33.09	1.75
523	GLU-35	126.19	8.71	57.19	4.67	31.66	2.01/2.05
524	LYS-36	123.74	8.84	54.11	4.72	36.05	1.7
525	GLU-37	117.33	8.48	55.21	5.55	33.48	1.81/1.75

526	ALA-38	120.83	8.27	51.59	4.42	23.35	1.17
527	TYR-39	119.28	8.33	60.04	4.63	39.07	3.21/2.57
528	SER-40	114.17	8.35	58	4.5	65.3	3.7/3.61
529	ASP-41	125.04	8.82	55.94	4.51	41.55	2.63/2.46
530	ASP-42	114.44	8.07	54.62	4.65	43.05	2.5
531	VAL-43	121.1	7.87	60.27	4.38	33.47	1.81
532	PRO-44			63.55	4.36	32.56	2.21
533	GLU-45	119.18	8.54	57.89	3.65	29.67	1.8/2.04
534	GLY-46	111.79	9.18	45.19	4.33/3.72		
535	GLN-47	115.28	7.67	54.25	5.21	31.1	1.96
536	VAL-48	119.24	8.51	64.85	3.48	32.64	2.2
537	VAL-49	124.65	9.23	63.5	3.94	34.75	1.51
538	LYS-50	112.98	7.18	55.09	4.56	35.17	1.79/1.66
539	GLN-51	114.83	8.26	53.43	5.39	34.78	1.64/1.53
540	LYS-52	121.76	8.77	54.61	4.17	36.7	1.54/1.64
541	PRO-53			62.77	4.52	34.93	2.05/2.48
542	ALA-54	115.84	7.98	51.97	4	20.04	1.32
543	ALA-55	122.14	8.77	53.76	3.44	18.2	1.14
544	GLY-56	109.35	9.13	44.98	4.25/3.29		
545	THR-57	115.96	7.49	64.08	3.99	69.9	3.77
546	ALA-58	127.76	8.48	51.79	4.96	18.8	1.22
547	VAL-59	117.84	9.09	59.37	4.57	35.44	2.12
548	LYS-60	118.63	7.81	53.22	4.82	33.75	1.7/1.42
549	PRO-61			64.53	3.96	31.82	2.3
550	GLY-62	110.56	8.49	44.93	4.35/3.49		
551	ASN-63	116.83	7.9	53.62	4.57	40.06	2.81/2.75
552	GLU-64	116.94	8.37	55.56	4.81	31.84	1.86/1.72
553	VAL-65	121.31	8.75	61.71	4.41	34.54	1.81
554	GLU-66	126.44	9.11	54.61	5.16	31.36	1.91/1.86
555	VAL-67	128.21	8.75	61.74	4.15	33.65	1.55
556	THR-68	119.86	8.11	61.11	4.3	69.11	3.63
557	PHE-69	126.33	9.16	57.77	4.78	41.82	2.93/2.8
558	SER-70	111.89	8.87	56.76	4.81	64.61	4.24/3.57
559	LEU-71	128.96	8.83	54.37	4.45	44.2	1.39/0.81
560	GLY-72	108.49	8.88	44.7	4.47/3.84		
561	PRO-73			63.28	4.43	32.39	2.28
562	GLU-74	126.77	8.46	59.01	3.11	30.2	1.42/1.25

Los δ del ^{15}N y del ^1H amídicos fueron determinados mediante experimentos HSQC, Los δ del $^{13}\text{C}_\alpha$ con experimento HNCA. Los δ del $^{13}\text{C}_\beta$ con HNCACB. Los δ del ^1H se determinaron mediante experimentos 3D-TOCSY-HSQC, 3D-NOESY-HSQC y 2D-NOESY.

Así, se han asignado los δ del 91% de los ^1H , y de todos los ^{15}N de la cadena peptídica, excepto los correspondientes a las cinco prolinas y los dos aminoácidos iniciales y los ^{15}N de las cadenas laterales básicas (80% del total de ^{15}N asignados) y de un 99% de los δ de los ^{13}C alifáticos (ver tabla 3.6) (65% del total de ^{13}C asignados), mientras que los ^{13}C aromáticos y carbonílicos no se asignaron. Estos δ asignados se utilizan en el programa CYANA para el cálculo de la estructura, como se describe a continuación.

3.2.4 Modelización molecular de la estructura 3D del dominio PASTA-3 recombinante de PreK de *B. subtilis*, mediante el programa CYANA (2.0)

Para el cálculo de la estructura se utilizó el programa CYANA 2.0, que usa una combinación de cálculo de estructuras por geometría de distancias, dinámica molecular en ángulos de torsión y asignación automática de los espectros NOESY (<http://guentert.gsc.riken.go.jp/software/Cyana.html>) [113-114]. Este protocolo parte de la asignación de los desplazamientos químicos de la mayoría de los núcleos de la proteína y de las listas de todas las señales NOE presentes en el espectro NOESY que se recopilaban mediante un proceso de etiquetado e integración de los picos de cruce en los espectros NOESY usando el programa CARRA. Con estos archivos de entrada, el listado de asignaciones de desplazamientos químicos a los protones de la proteína y el listado de coordenadas de desplazamiento químico e integrales de los picos de cruce de los espectros NOESY, el programa CYANA realiza un proceso iterativo de siete ciclos de cálculo de estructuras asignando las señales NOE a determinados pares de núcleos. Después de cada ciclo el programa usa cada estructura generada para recalcular un nuevo conjunto de restricciones de distancia y lo compara con los NOEs experimentales corrigiendo su asignación. El proceso iterativo transcurre hasta que se asignan más del 95% de las señales NOE, de forma que la gran mayoría de las restricciones de distancia derivadas de las estructuras son compatibles con la asignación de los picos de cruce NOE. El primer ciclo de cálculo parte de un conjunto de estructuras generadas al azar y en los siguientes ciclos se utilizan los procedentes del ciclo anterior. Las restricciones se incorporaron en todos los ciclos del cálculo de estructuras. Se calcularon un total de 100 estructuras, considerándose las 20 mejores, es decir, las que tienen un valor de la función blanco (ver *Materiales y métodos*), como las representativas de la estructura del dominio PASTA. En la tabla 3.7 se muestran, esquemáticamente, las restricciones utilizadas para generar el conjunto definitivo de las 20 estructuras del péptido. El valor

de RMSD es de 0.46Å para la cadena principal y de 0.96Å para los átomos pesados (tabla 3.7). En la figura 3.38, panel A, se presenta la superposición de las 20 estructuras finales generados con programa CYANA, para el dominio PASTA-3 recombinante (residuos 5-74)

Tabla 3.7 Datos estadísticos de las 20 estructuras de menos energía generadas mediante el programa CYANA 2,0 en cada ciclo para el cálculo de la estructura 3D del dominio PASTA-3 recombinante de PrkK de *B.subtilis*.

	Ciclo 1	Ciclo 2	Ciclo 3	Ciclo 4	Ciclo 5	Ciclo 6	Ciclo 7	Final
Picos Seleccionados	2834	2834	2834	2834	2834	2834	2834	
Picos Asignados	2660	2708	2685	2697	2690	2687	2678	
Picos No asignados	174	126	149	137	144	147	156	
Picos con asignación diagonal	359	359	359	359	359	359	359	
Los límites superiores de distancia								
Total	1067	942	904	904	887	883	932	968
Corta distancia, $i-j \leq 1$	720	615	594	575	560	556	549	566
Media distancia, $1 < i-j < 5$	229	181	97	99	92	91	93	100
Larga distancia, $i-j \geq 5$	118	146	213	230	235	236	290	302
Restricciones por residuo	4.08	1.90	1.31	1.30	1.24	1.20	1.00	1.00
Promedio del valor de función objetivo	15.07	10.06	16.11	2.35	1.21	0.68	0.47	0.24
RMSD (residuos 5-74)								
RMSD promedio de la cadena principal	3.14	1.81	0.77	0.56	0.54	0.50	0.46	0.46
RMSD promedio de átomos pesados	3.77	2.33	1.31	1.07	1.03	0.99	0.97	0.96

Otra forma de comprobar la calidad de la estructura es el diagrama de Ramachandran, en el que se representa el ángulo ϕ (Φ) frente al ψ (Ψ) para cada residuo de aminoácido. En la figura 3.38 B se muestra el diagrama de Ramachandran para la estructura 3D de menor energía del dominio PASTA-3 de PrkC, en el que se han distinguido distintas zonas según su grado de permisividad, mediante un código de colores: desde blanco para las regiones no permitidas, hasta azul intenso para las regiones más favorables para las cadenas polipeptídicas. Así, el diagrama de Ramachandran nos indica que todos los residuos se encuentran en regiones permitidas o permitidas adicionalmente.

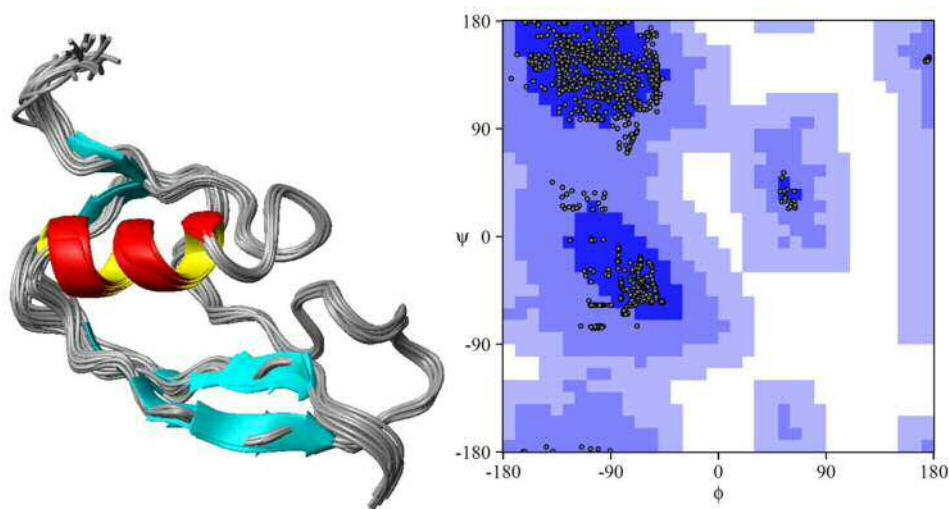


Figura 3.38 **A)** Modelo de cintas y flechas del dominio PASTA-3 de PrkC de *B. subtilis* donde se muestra la superposición de las 20 estructuras finales del dominio recombinante (residuos 5-74). La imagen fue generada con el programa MolMol [115]. Se usa la convención: cintas, hélice α ; y flechas, lámina β . **B)** Representación de Ramachandran de la estructura de PASTA-3 de PrkC de *B. subtilis* de menor energía excluyendo los primeros 3 aminoácidos. Se muestra para cada aminoácido el par de ángulos diedros ϕ y ψ correspondiente. La escala de colores ilustra el grado de permisibilidad: azul intenso, regiones más favorables; azul ligeramente claro, regiones permitidas adicionalmente; azul claro, zonas permitidas generosamente; y blanco, regiones no permitidas.

En la serie de estructuras de obtenidas mediante el programa CYANA, aún sin realizar una posterior optimización mediante dinámica molecular, se empiezan a definir algunos de los principales elementos de estructura secundaria descritos en otros dominios PASTA: una α hélice seguida de hasta 3 láminas β (Figura 3.39).

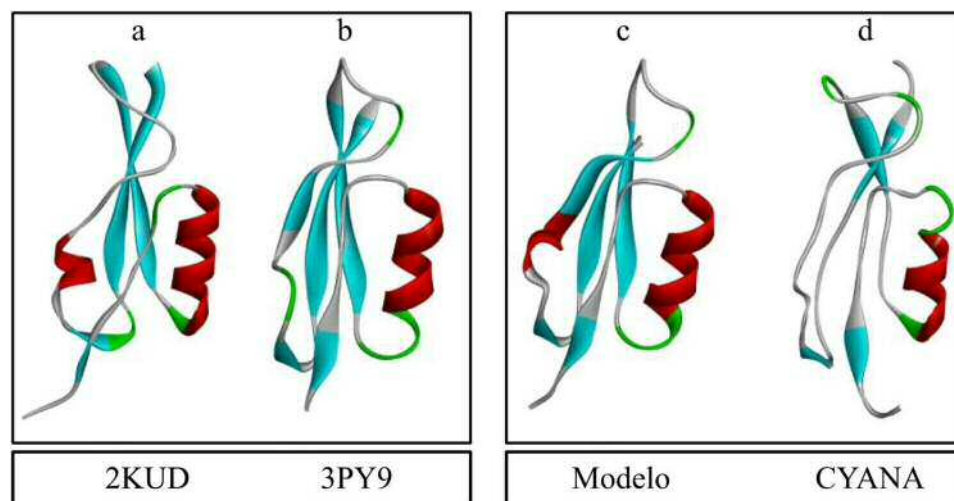


Figura 3.39 Estructuras 3D de varios dominios PASTA: **a)** dominios PASTA-2 de PknB de *Mycobacterium tuberculosis* (código PDB 2KUD) determinada mediante RMN, **b)** dominio PASTA-3 de PrkC de *S. aureus* (código PDB 3PY9) determinada mediante cristalografía de rayos X, **c)** modelo de homología obtenido con la estructura de 3PY9, utilizando el programa SwissModel (<http://swissmodel.expasy.org/> [116]) y **d)** dominio PASTA-3 recombinante de PrkC de *B. subtilis*, determinado mediante RMA en la presente tesis. Se usan los colores: rojo para la hélice α ; y azul turquesa para la lámina β .

A primera vista, el plegamiento global que se observa en la estructura no refinada obtenida para el dominio PASTA-3 recombinante de PrkC de *B. subtilis* (Figura 3.39) es análogo a la que se observa en otros dominios PASTA cuyas estructuras 3D han sido determinadas (Figura 3.39). En el *Protein Data Bank* se pueden encontrar hasta 17 estructuras 3D de dominios PASTA: 5 de ellas resueltas por RMN (2MGV, 2KUD, 2KUE, 2KUF y 2KUI) y las 12 restantes, por cristalografía de rayos-X (3OUV, 3PY9, 2ZC3, 2ZC4, 2Z2L, 2Z2M, 1RP5, 1PYY, 1K25, 1QME, 1QMF y 1PMD). En todos los casos no se observan las 3 laminas- β completamente definidas las 3 laminas β . Por ejemplo, en la estructura del dominio PASTA-2 de PknB de *Mycobacterium tuberculosis* (código PDB 2KUD) no está bien definida la lamina- β 2 (Figura 3.39-a). Sin embargo, sí se observa una lamina- β incipiente (β 0 y β 2') entre el extremo amino terminal y el "loop" entre las laminas β 2 y β 3. Respecto a la hélice α , en la estructura obtenida no llegan a definirse completamente las dos vueltas completas de la hélice que se observan en las otras estructuras.

A modo de comparación, se muestra un modelo estructural del dominio PASTA-3 de PrkC de *B. subtilis* obtenido por homología a partir de la estructura del dominio

PASTA-3 de la PrkC de *S. aureus*, con la que comparte un 35% de identidad secuencial. En este modelo, sí se observan todos los elementos de estructura secundaria asociados a los dominios PASTA por lo que se utilizó como plantilla estructural para construir el modelo, calculado en el servidor SwissModel, (<http://swissmodel.expasy.org>) [116].

3.2.5 Estudios de interacción entre el dominio PASTA-3 recombinante de PrkC de *B. subtilis* y varios carbohidratos

Una vez asignado el espectro de la proteína, aún sin tener una estructura totalmente refinada, es posible estudiar las interacciones con potenciales ligandos aplicando la estrategia de observación de las señales del receptor sobre los espectros ^{15}N -HSQC (Figura 3.37). Este método es muy sensible y permite detectar perturbaciones en el entorno de las amidas de la cadena principal del dominio PASTA-3, debidas a la unión al ligando.

Se realizaron experimentos ^1H - ^{15}N HSQC para monitorizar y determinar cambios en los δ químicos del dominio PASTA-3 de PrkC de *B. subtilis* durante la interacción con cantidades crecientes del muropéptido extraído de la misma bacteria, que a su vez fue analizado mediante experimentos 2D de RMN para comprobar la presencia de los residuos DAP (figura 3.40). También se realizaron experimentos con ampicilina (antibiótico β -lactámico). Estos compuestos han sido previamente descritos como ligandos de dominio PASTA [8, 92-93].

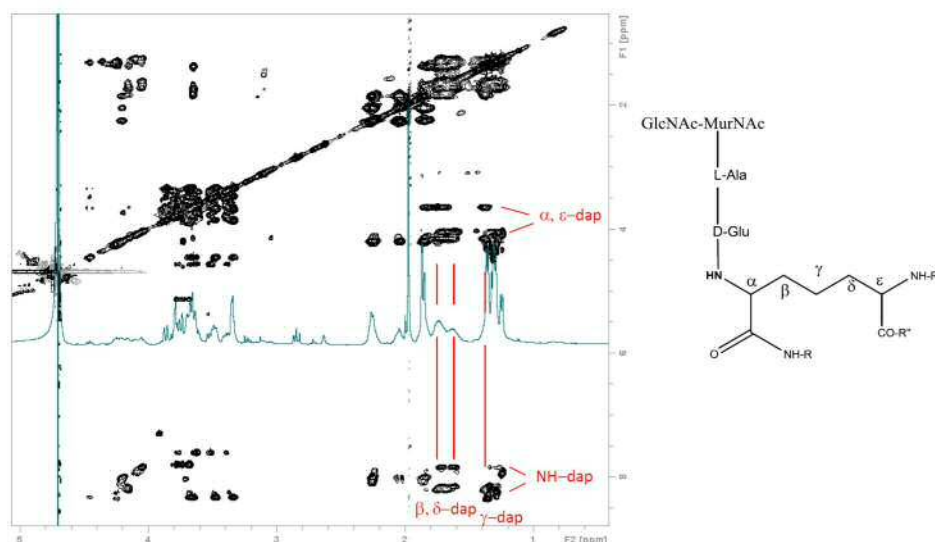


Figura 3.40 Espectro TOCSY a 60 ms de la fracción de muropéptidos aislados de peptidoglicano de *B. subtilis* que contienen residuos DAP en su estructura. En el espectro se indican los picos de correlación entre los protones amídicos del DAP y los protones alifáticos. Adquirido en un espectrómetro Bruker de 600 MHz a 298K.

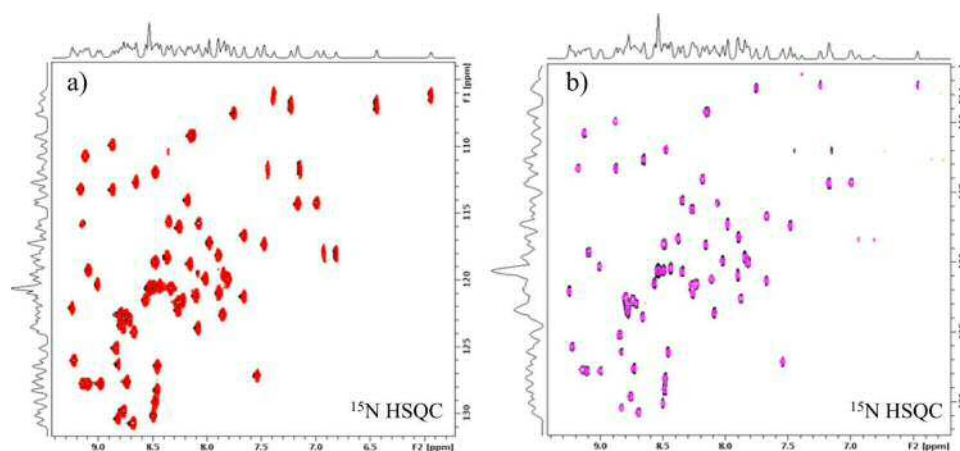


Figura 3.41 Espectros ^1H - ^{15}N -HSQC adquiridos en 600MHz del dominio PASTA en presencia de potenciales ligandos. **Panel a)** Puntos inicial (ausencia de ligando, negro) y final (relación 20:1, rojo) de la titulación con un muropéptido extraído de *B. subtilis*. **Panel b)** Puntos inicial (ausencia de ligando, negro) y final (relación 10:1, rojo) de la titulación con el betalactámico ampicilina.

En la figura 3.41 se muestran los espectros ^1H - ^{15}N HSQC obtenidos del dominio PASTA-3 (0,4 mM), tanto en ausencia de muropéptido como en presencia de un exceso de éste último de 20:1. Análogamente, se realizó un experimento similar de titulación

con la ampicilina sobre una muestra del dominio PASTA (0,2 mM) hasta un exceso 10:1 del antibiótico.

En ninguno de los dos casos (muropéptido y ampicilina) se observaron desplazamientos significativos en los picos de correlación H-N correspondientes a los enlaces peptídicos que pudieran informar de la unión de estos ligandos al dominio PASTA-3.

En estos estudios, el GMDP y el tripéptido N-Acetil-L-Lisina-D-Alanina-D-Alanina, ambos se utilizaron como controles negativos, ya que no contienen el residuo de DAP. Como era de esperar, con ninguno de estos compuestos se observó una perturbación significativa en los espectros ^{15}N -HSQC del dominio PASTA-3.

En resumen, no se ha detectado la interacciones de un único dominio PASTA-3 de PrkC de *B. subtilis* y muropéptido de tipo DAP derivado del peptidoglicano de la propia bacteria. Como se ha mencionado, el DAP se ha propuesto como elemento esencial para el reconocimiento y la selectividad de PrkC de *B. subtilis* hacia el PGN. Además, se ha demostrado, mediante experimentos de STD-RMN, que dichos fragmentos de PGN (muropéptidos) son capaces de unirse al segmento extracelular completo de PrkC, que contiene 3 dominios PASTA consecutivos [20].

De manera análoga a lo discutido para el caso de los dominios LysM, queda por confirmar estructuralmente si en la interacción entre PrkC y PGN intervienen más de uno de los dominios PASTA de la proteína e incluso si se producen procesos de oligomerización mediados por ligando. Recientemente, un estudio publicado en abril 2014 [117], los autores han demostrado que un único dominio PASTA de una proteína PBP de *Micobacterium tuberculosis* (código PDB: 2MGV) no es capaz de interactuar con los ligandos descritos para la proteína completa: antibióticos betalactámicos, muropéptidos o peptidoglicano polimérico [104, 109-110, 117].

3.3 CONCLUSIONES

Se han asignado los δ del dominio PASTA-3 recombinante de PrkC de *B. subtilis* mediante la aplicación de experimentos multidimensionales homo- y heteronucleares de RMN lo que ha permitido el cálculo de una estructura 3D no refinada en base a los datos experimentales de NOE que ha confirmado la presencia de los elementos de estructura secundaria y tipo de plegamiento característicos de los dominios PASTA.

Según los experimentos de interacción por RMN, el dominio PASTA-3 recombinante de PrkC de *B. subtilis* parece no interacciona con los ligandos descritos anteriormente para el segmento extracelular completo de la proteína PrkC: ampicilina y muropéptidos de tipo DAP derivados del PGN de *B. subtilis*.

4 CONCLUSIONES PARA EL CAPITULO 3

En cada uno de los apartados de este capítulo, que recogen los resultados experimentales, se han incluido las conclusiones concretas de los sistemas abordados y las técnicas utilizadas. A modo de resumen, incluimos aquí una serie de conclusiones generales que pueden deducir de este trabajo centrado en las interacciones de dominios de proteínas con carbohidratos derivados de GlcNAc, que se han estudiado por RMN.

La unión del GMDP a la WGA usando como modelo el dominio B (el segundo dominio tipo heveína en la secuencia de la WGA) se establece según el modo de unión donde el anillo de GlcNAc se sitúa sobre la Tyr 66, donde la parte peptídica del GMDP queda esencialmente expuesta al disolvente, de acuerdo con los bajos índices de saturación observados para los residuos de alanina y isoglutamina (< 4%). El grupo metilo del resto de ácido láctico en la posición 3 del ácido murámico (CH₃-Lac-Mβ) apunta hacia la Tyr 64 y la superficie de la proteína, justificándose su significativa saturación (11%).

Se ha confirmado experimentalmente por RMN que el dominio LysM derivado de la transglicosilasa lítica de *E. coli* interacciona con el peptidoglicano, usando como modelo el péptido recombinante LysM-CAM-1 que incluye la secuencia LysM 389-452 de la proteína MltD con dos mutaciones S388C y N453C. Sin embargo, que no se observaron interacciones con el mínimo elemento disacárido, el GMDP, que se repite en la estructura del peptidoglicano.

Se han asignado los δ del dominio PASTA-3 recombinante de PrkC de *B. subtilis* mediante la aplicación de experimentos multidimensionales homo- y heteronucleares de RMN lo que ha permitido el cálculo de una estructura 3D no refinada. Según los experimentos de interacción por RMN, el dominio PASTA-3 parece no interacciona con los ligandos descritos anteriormente para el segmento extracelular completo de la proteína PrkC: ampicilina y muropéptidos de tipo DAP derivados del PGN de *B. subtilis*.

5 REFERENCIAS BIBLIOGRAFICAS

1. A. Kobata, *Structures and functions of the sugar chains of glycoproteins*. Eur J Biochem, 1992. **209**(2): p. 483-501.
2. R.A. Dwek, *Glycobiology: Toward Understanding the Function of Sugars*. Chem Rev, 1996. **96**(2): p. 683-720.
3. J.V. Cullimore, R. Ranjeva y J.-J. Bono, *Perception of lipo-chitooligosaccharidic Nod factors in legumes*. Trends in Plant Science, 2001. **6**(1): p. 24-30.
4. A. Ayoub, H. Causse, E.J.M. Van Damme, W.J. Peumans, Y. Bourne, C. Cambillau y P. Rougé, *Interactions of plant lectins with the components of the bacterial cell wall peptidoglycan*. Biochemical Systematics and Ecology, 1994. **22**(2): p. 153-159.
5. J.L. Asensio, F.J. Canada, H.C. Siebert, J. Laynez, A. Poveda, P.M. Nieto, U.M. Soedjanaamadja, H.J. Gabius y J. Jimenez-Barbero, *Structural basis for chitin recognition by defense proteins: GlcNAc residues are bound in a multivalent fashion by extended binding sites in hevein domains*. Chem Biol, 2000. **7**(7): p. 529-543.
6. R. Lotan, N. Sharon y D. Mirelman, *Interaction of wheat-germ agglutinin with bacterial cells and cell-wall polymers*. Eur J Biochem, 1975. **55**(1): p. 257-262.
7. G. Buist, A. Steen, J. Kok y O.P. Kuipers, *LysM, a widely distributed protein motif for binding to (peptido)glycans*. Molecular Microbiology, 2008. **68**(4): p. 838-847.
8. C. Yeats, R.D. Finn y A. Bateman, *The PASTA domain: a beta-lactam-binding domain*. Trends Biochem Sci, 2002. **27**(9): p. 438.
9. E. Limpens, C. Franken, P. Smit, J. Willemse, T. Bisseling y R. Geurts, *LysM domain receptor kinases regulating rhizobial Nod factor-induced infection*. Science, 2003. **302**(5645): p. 630-633.
10. S. Onaga y T. Taira, *A new type of plant chitinase containing LysM domains from a fern (Pteris ryukyuensis): roles of LysM domains in chitin binding and antifungal activity*. Glycobiology, 2008. **18**(5): p. 414-423.
11. J. Jimenez-Barbero, F. Javier Canada, J.L. Asensio, N. Aboitiz, P. Vidal, A. Canales, P. Groves, H.J. Gabius y H.C. Siebert, *Hevein domains: an attractive model to study carbohydrate-protein interactions at atomic resolution*. Advances in carbohydrate chemistry and biochemistry, 2006. **60**: p. 303-354.
12. A.M. Benko-Iseppon, S.L. Galdino, T. Calsa, E.A. Kido, A. Tossi, L.C. Belarmino y S. Crovella, *Overview on Plant Antimicrobial Peptides*. Current Protein & Peptide Science, 2010. **11**(3): p. 181-188.
13. R.R. Shukurov, V.D. Voblikova, A.K. Nikonorova, R.A. Komakhin, V.V. Komakhina, T.A. Egorov, E.V. Grishin y A.V. Babakov, *Transformation of tobacco and Arabidopsis plants with Stellaria media genes encoding novel hevein-like peptides increases their resistance to fungal pathogens*. Transgenic research, 2011.
14. D.V. Kurek, S.A. Lopatin y V.P. Varlamov, *Prospects of application of the chitin-binding domains to isolation and purification of recombinant proteins by affinity chromatography*. Appl. Biochem. Microbiol., 2009. **45**(1): p. 1-8.
15. R. Hammami, J. Ben Hamida, G. Vergoten y I. Fliss, *PhytAMP: a database dedicated to antimicrobial plant peptides*. Nucleic acids research, 2009. **37**(Database issue): p. D963-968.
16. K. Laroy y G. Weeks, *Inhibition of Dictyostelium discoideum differentiation in monolayers in vitro by endogenous and exogenous lectins*. J Cell Sci, 1983. **59**: p. 203-212.
17. D. Mirelman, E. Galun, N. Sharon y R. Lotan, *Inhibition of fungal growth by wheat germ agglutinin*. Nature, 1975. **256**(5516): p. 414-416.
18. J.L. Asensio, F.J. Cañada, H.-C. Siebert, J. Laynez, A. Poveda, P.M. Nieto, U.M. Soedjanaamadja, H.-J. Gabius y J. Jiménez-Barbero, *Structural basis for chitin*

- recognition by defense proteins: GlcNAc residues are bound in a multivalent fashion by extended binding sites in hevein domains.* Chemistry & Biology, 2000. **7**(7): p. 529-543.
19. J.C. Aub, B.H. Sanford y L.H. Wang, *Reactions of normal and leukemic cell surfaces to a wheat germ agglutinin.* Proc Natl Acad Sci U S A, 1965. **54**(2): p. 400-402.
 20. K. Harata y M. Muraki, *X-ray structure of turkey egg lysozyme complex with di-N-acetyl-chitobiose. Recognition and binding of alpha-anomeric form.* Acta Crystallogr D Biol Crystallogr, 1995. **51**(Pt 5): p. 718-724.
 21. C.S. Wright, *Crystallographic elucidation of the saccharide binding mode in wheat germ agglutinin and its biological significance.* J Mol Biol, 1980. **141**(3): p. 267-291.
 22. C.S. Wright, *Comparison of the refined crystal structures of two wheat germ isolectins.* J Mol Biol, 1989. **209**(3): p. 475-487.
 23. C.S. Wright, *2.2 A resolution structure analysis of two refined N-acetylneuraminyllactose--wheat germ agglutinin isolectin complexes.* J Mol Biol, 1990. **215**(4): p. 635-651.
 24. C.S. Wright y J. Jaeger, *Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycophorin A.* J Mol Biol, 1993. **232**(2): p. 620-638.
 25. C.S. Wright, *Structural comparison of the two distinct sugar binding sites in wheat germ agglutinin isolectin II.* J Mol Biol, 1984. **178**(1): p. 91-104.
 26. M. Muraki, M. Ishimura y K. Harata, *Interactions of wheat-germ agglutinin with GlcNAc beta 1,6Gal sequence.* Biochim Biophys Acta, 2002. **1569**(1-3): p. 10-20.
 27. D. Schwefel, C. Maierhofer, J.G. Beck, S. Seeberger, K. Diederichs, H.M. Moller, W. Welte y V. Wittmann, *Structural basis of multivalent binding to wheat germ agglutinin.* J Am Chem Soc, 2010. **132**(25): p. 8704-8719.
 28. J.J. Hernández-Gay, A. Ardá, S. Eller, S. Mezzato, B.R. Leeftang, C. Unverzagt, F.J. Cañada y J. Jiménez-Barbero, *Insights into the Dynamics and Molecular Recognition Features of Glycopeptides by Protein Receptors: The 3D Solution Structure of Hevein Bound to the Trisaccharide Core of N-Glycoproteins.* Chemistry – A European Journal, 2010. **16**(35): p. 10715-10726.
 29. J.L. Asensio, F.J. Canada, M. Bruix, A. Rodríguez-Romero y J. Jiménez-Barbero, *The interaction of hevein with N-acetylglucosamine-containing oligosaccharides. Solution structure of hevein complexed to chitobiose.* Eur J Biochem, 1995. **230**(2): p. 621-633.
 30. J.F. Espinosa, J.L. Asensio, J.L. García, J. Laynez, M. Bruix, C. Wright, H.C. Siebert, H.J. Gabius, F.J. Canada y J. Jiménez-Barbero, *NMR investigations of protein-carbohydrate interactions binding studies and refined three-dimensional solution structure of the complex between the B domain of wheat germ agglutinin and N,N', N''-triacylchitotriose.* Eur J Biochem, 2000. **267**(13): p. 3965-3978.
 31. S. Yokoyama, Y. Iida, Y. Kawasaki, Y. Minami, K. Watanabe y F. Yagi, *The chitin-binding capability of Cy-AMP1 from cycad is essential to antifungal activity.* J Pept Sci, 2009. **15**(7): p. 492-497.
 32. P.V. Dubovskii, A.A. Vassilevski, A.A. Slavokhotova, T.I. Odintsova, E.V. Grishin, T.A. Egorov y A.S. Arseniev, *Solution structure of a defense peptide from wheat with a 10-cysteine motif.* Biochemical and biophysical research communications, 2011. **411**(1): p. 14-18.
 33. R.H. Huang, Y. Xiang, G.Z. Tu, Y. Zhang y D.C. Wang, *Solution structure of Eucommia antifungal peptide: a novel structural model distinct with a five-disulfide motif.* Biochemistry, 2004. **43**(20): p. 6005-6012.
 34. M. Wagner, *Interaction of wheat-germ agglutinin with streptococci and streptococcal cell wall polymers.* Immunobiology, 1979. **156**(1-2): p. 57-64.
 35. R.K. Sizemore, J.J. Caldwell y A.S. Kendrick, *Alternate gram staining technique using a fluorescent lectin.* Applied and environmental microbiology, 1990. **56**(7): p. 2245-2247.

36. M. Mayer y B. Meyer, *Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor*. J Am Chem Soc, 2001. **123**(25): p. 6108-6117.
37. C.S. Wright, *Crystal structure of a wheat germ agglutinin/glycophorin-sialoglycopeptide receptor complex. Structural basis for cooperative lectin-cell binding*. J Biol Chem, 1992. **267**(20): p. 14345-14352.
38. M.I. Chávez, C. Andreu, P. Vidal, N. Aboitiz, F. Freire, P. Groves, J.L. Asensio, G. Asensio, M. Muraki, F.J. Cañada y J. Jiménez-Barbero, *On the Importance of Carbohydrate-Aromatic Interactions for the Molecular Recognition of Oligosaccharides by Proteins: NMR Studies of the Structure and Binding Affinity of AcAMP2-like Peptides with Non-Natural Naphthyl and Fluoroaromatic Residues*. Chemistry – A European Journal, 2005. **11**(23): p. 7060-7074.
39. N. Aboitiz, M. Vila-Perelló, P. Groves, J.L. Asensio, D. Andreu, F.J. Cañada y J. Jiménez-Barbero, *NMR and Modeling Studies of Protein-Carbohydrate Interactions: Synthesis, Three-Dimensional Structure, and Recognition Properties of a Minimum Hevein Domain with Binding Affinity for Chitoooligosaccharides*. ChemBioChem, 2004. **5**(9): p. 1245-1255.
40. J.L. Asensio, F.J. Canada, M. Bruix, C. Gonzalez, N. Khiar, A. Rodriguez-Romero y J. Jimenez-Barbero, *NMR investigations of protein-carbohydrate interactions: refined three-dimensional structure of the complex between hevein and methyl beta-chitobioside*. Glycobiology, 1998. **8**(6): p. 569-577.
41. G. Colombo, M. Meli, J. Cañada, J.L. Asensio y J. Jiménez-Barbero, *Toward the understanding of the structure and dynamics of protein-carbohydrate interactions: molecular dynamics studies of the complexes between hevein and oligosaccharidic ligands*. Carbohydrate Research, 2004. **339**(5): p. 985-994.
42. A.D. Bain y G.J. Duns, *A unified approach to dynamic NMR based on a physical interpretation of the transition probability*. Canadian Journal of Chemistry, 1996. **74**(6): p. 819-824.
43. F.A. Saul, P. Rovira, G. Boulot, E.J. Damme, W.J. Peumans, P. Truffa-Bachi y G.A. Bentley, *Crystal structure of Urtica dioica agglutinin, a superantigen presented by MHC molecules of class I and class II*. Structure, 2000. **8**(6): p. 593-603.
44. E.J. van Asselt, K.H. Kalk y B.W. Dijkstra, *Crystallographic Studies of the Interactions of Escherichia coli Lytic Transglycosylase Slt35 with Peptidoglycan†*. Biochemistry, 2000. **39**(8): p. 1924-1934.
45. H.C. Siebert, C.W. von der Lieth, R. Kaptein, J.J. Beintema, K. Dijkstra, N. van Nuland, U.M. Soedjanaatmadja, A. Rice, J.F. Vliegthart, C.S. Wright y H.J. Gabius, *Role of aromatic amino acids in carbohydrate binding of plant lectins: laser photo chemically induced dynamic nuclear polarization study of hevein domain-containing lectins*. Proteins, 1997. **28**(2): p. 268-284.
46. G.E.D. Oldroyd y S. Robatzek, *The broad spectrum of plant associations with other organisms*. Current Opinion in Plant Biology, 2011. **14**(4): p. 347-350.
47. H. Kaku, Y. Nishizawa, N. Ishii-Minami, C. Akimoto-Tomiyama, N. Dohmae, K. Takio, E. Minami y N. Shibuya, *Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor*. Proceedings of the National Academy of Sciences, 2006. **103**(29): p. 11086-11091.
48. K. Kishimoto, Y. Kouzai, H. Kaku, N. Shibuya, E. Minami y Y. Nishizawa, *Perception of the chitin oligosaccharides contributes to disease resistance to blast fungus Magnaporthe oryzae in rice*. The Plant Journal, 2010. **64**(2): p. 343-354.
49. A. Miya, P. Albert, T. Shinya, Y. Desaki, K. Ichimura, K. Shirasu, Y. Narusaka, N. Kawakami, H. Kaku y N. Shibuya, *CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis*. Proceedings of the National Academy of Sciences, 2007. **104**(49): p. 19613-19618.

50. T. Shimizu, T. Nakano, D. Takamizawa, Y. Desaki, N. Ishii-Minami, Y. Nishizawa, E. Minami, K. Okada, H. Yamane, H. Kaku y N. Shibuya, *Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice*. The Plant Journal, 2010. **64**(2): p. 204-214.
51. R. Willmann, H.M. Lajunen, G. Erbs, M.-A. Newman, D. Kolb, K. Tsuda, F. Katagiri, J. Fliegmann, J.-J. Bono, J.V. Cullimore, A.K. Jehle, F. Götz, A. Kulik, A. Molinaro, V. Lipka, A.A. Gust y T. Nürnberger, *Arabidopsis lysin-motif proteins LYM1 LYM3 CERK1 mediate bacterial peptidoglycan sensing and immunity to bacterial infection*. Proceedings of the National Academy of Sciences, 2011. **108**(49): p. 19824-19829.
52. K.H. Schleifer y O. Kandler, *Peptidoglycan types of bacterial cell walls and their taxonomic implications*. Bacteriological reviews, 1972. **36**(4): p. 407-477.
53. A. Steen, G. Buist, K.J. Leenhouts, M. El Khatibi, F. Grijpstra, A.L. Zomer, G. Venema, O.P. Kuipers y J. Kok, *Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents*. Journal of Biological Chemistry, 2003. **278**(26): p. 23874-23881.
54. G. Andre, K. Leenhouts, P. Hols y Y.F. Dufrene, *Detection and localization of single LysM-peptidoglycan interactions*. J Bacteriol, 2008. **190**(21): p. 7079-7086.
55. A. Bateman y M. Bycroft, *The structure of a LysM domain from E-coli membrane-bound lytic murein transglycosylase D (MltD)*. Journal of Molecular Biology, 2000. **299**(4): p. 1113-1119.
56. C. Zipfel, *Pattern-recognition receptors in plant innate immunity*. Current Opinion in Immunology, 2008. **20**(1): p. 10-16.
57. T. Boller y G. Felix, *A Renaissance of Elicitors: Perception of Microbe-Associated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors*. Annual Review of Plant Biology, 2009. **60**(1): p. 379-406.
58. M. Hayafune, R. Berisio, R. Marchetti, A. Silipo, M. Kayama, Y. Desaki, S. Arima, F. Squeglia, A. Ruggiero, K. Tokuyasu, A. Molinaro, H. Kaku y N. Shibuya, *Chitin-induced activation of immune signaling by the rice receptor CEBiP relies on a unique sandwich-type dimerization*. Proceedings of the National Academy of Sciences, 2014. **111**(3): p. E404-E413.
59. Y. Li, Y.G. Lu, Y. Shi, L. Wu, Y.J. Xu, F. Huang, X.Y. Guo, Y. Zhang, J. Fan, J.Q. Zhao, H.Y. Zhang, P.Z. Xu, J.M. Zhou, X.J. Wu, P.R. Wang y W.M. Wang, *Multiple rice microRNAs are involved in immunity against the blast fungus Magnaporthe oryzae*. Plant Physiol, 2014. **164**(2): p. 1077-1092.
60. J. Fliegmann, S. Uhlenbroich, T. Shinya, Y. Martinez, B. Lefebvre, N. Shibuya y J.J. Bono, *Biochemical and phylogenetic analysis of CEBiP-like LysM domain-containing extracellular proteins in higher plants*. Plant Physiol Biochem, 2011. **49**(7): p. 709-720.
61. J. Wan, X.C. Zhang, D. Neece, K.M. Ramonell, S. Clough, S.Y. Kim, M.G. Stacey y G. Stacey, *A LysM receptor-like kinase plays a critical role in chitin signaling and fungal resistance in Arabidopsis*. Plant Cell, 2008. **20**(2): p. 471-481.
62. T. Liu, Z. Liu, C. Song, Y. Hu, Z. Han, J. She, F. Fan, J. Wang, C. Jin, J. Chang, J.M. Zhou y J. Chai, *Chitin-induced dimerization activates a plant immune receptor*. Science, 2012. **336**(6085): p. 1160-1164.
63. A. Kombrink, A. Sanchez-Vallet y B.P. Thomma, *The role of chitin detection in plant-pathogen interactions*. Microbes Infect, 2011. **13**(14-15): p. 1168-1176.
64. K.J. Garvey, M.S. Saedi y J. Ito, *Nucleotide sequence of Bacillus phage phi 29 genes 14 and 15: homology of gene 15 with other phage lysozymes*. Nucleic Acids Res, 1986. **14**(24): p. 10001-10008.
65. M.S. Saedi, K.J. Garvey y J. Ito, *Cloning and purification of a unique lysozyme produced by Bacillus phage phi 29*. Proc Natl Acad Sci U S A, 1987. **84**(4): p. 955-958.
66. D.M. Donovan, *Bacteriophage and peptidoglycan degrading enzymes with antimicrobial applications*. Recent Pat Biotechnol, 2007. **1**(2): p. 113-122.

67. S. Layec, J. Gerard, V. Legue, M.P. Chapot-Chartier, P. Courtin, F. Borges, B. Decaris y N. Leblond-Bourget, *The CHAP domain of Cse functions as an endopeptidase that acts at mature septa to promote Streptococcus thermophilus cell separation*. Mol Microbiol, 2009. **71**(5): p. 1205-1217.
68. A. Moll, S. Schlimpert, A. Briegel, G.J. Jensen y M. Thanbichler, *DipM, a new factor required for peptidoglycan remodelling during cell division in Caulobacter crescentus*. Mol Microbiol, 2010. **77**(1): p. 90-107.
69. T. Costa, A.L. Isidro, C.P. Moran, Jr. y A.O. Henriques, *Interaction between coat morphogenetic proteins SafA and SpoVID*. J Bacteriol, 2006. **188**(22): p. 7731-7741.
70. A. Sánchez-Vallet, R. Saleem-Batcha, A. Kombrink, G. Hansen, D.-J. Valkenburg, B.P.H.J. Thomma, J.R. Mesters y J. Greenberg, *Fungal effector Ecp6 outcompetes host immune receptor for chitin binding through intrachain LysM dimerization*. Vol. 2. 2013.
71. E.B. Madsen, L.H. Madsen, S. Radutoiu, M. Olbryt, M. Rakwalska, K. Szczygłowski, S. Sato, T. Kaneko, S. Tabata, N. Sandal y J. Stougaard, *A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals*. Nature, 2003. **425**(6958): p. 637-640.
72. J. Cullimore y J. Denarie, *Plant sciences. How legumes select their sweet talking symbionts*. Science, 2003. **302**(5645): p. 575-578.
73. C. Gough y J. Cullimore, *Lipo-chitoooligosaccharide Signaling in Endosymbiotic Plant-Microbe Interactions*. Molecular Plant-Microbe Interactions, 2011. **24**(8): p. 867-878.
74. R. Op den Camp, A. Streng, S. De Mita, Q. Cao, E. Polone, W. Liu, J.S.S. Ammiraju, D. Kudrna, R. Wing, A. Untergasser, T. Bisseling y R. Geurts, *LysM-Type Mycorrhizal Receptor Recruited for Rhizobium Symbiosis in Nonlegume Parasponia*. Science, 2011. **331**(6019): p. 909-912.
75. S. Radutoiu, L.H. Madsen, E.B. Madsen, H.H. Felle, Y. Umehara, M. Gronlund, S. Sato, Y. Nakamura, S. Tabata, N. Sandal y J. Stougaard, *Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases*. Nature, 2003. **425**(6958): p. 585-592.
76. R.F. Fisher y S.R. Long, *RHIZOBIUM - PLANT SIGNAL EXCHANGE*. Nature, 1992. **357**(6380): p. 655-660.
77. H.P. Spaink, *Root nodulation and infection factors produced by rhizobial bacteria*. Annu Rev Microbiol, 2000. **54**: p. 257-288.
78. S.R. Long, *Rhizobium symbiosis: nod factors in perspective*. Plant Cell, 1996. **8**(10): p. 1885-1898.
79. F. Debelle, C. Plazanet, P. Roche, C. Pujol, A. Savagnac, C. Rosenberg, J.C. Prome y J. Denarie, *The NodA proteins of Rhizobium meliloti and Rhizobium tropici specify the N-acylation of Nod factors by different fatty acids*. Mol Microbiol, 1996. **22**(2): p. 303-314.
80. X. Perret, C. Staehelin y W.J. Broughton, *Molecular basis of symbiotic promiscuity*. Microbiology and Molecular Biology Reviews, 2000. **64**(1): p. 180-+.
81. L. Mulder, B. Lefebvre, J. Cullimore y A. Imberty, *LysM domains of Medicago truncatula NFP protein involved in Nod factor perception. Glycosylation state, molecular modeling and docking of chitoooligosaccharides and Nod factors*. Glycobiology, 2006. **16**(9): p. 801-809.
82. J. Bielnicki, Y. Devedjiev, U. Derewenda, Z. Dauter, A. Joachimiak y Z.S. Derewenda, *B. subtilis ykuD protein at 2.0 Å resolution: Insights into the structure and function of a novel, ubiquitous family of bacterial enzymes*. Proteins: Structure, Function, and Bioinformatics, 2006. **62**(1): p. 144-151.
83. K.L. Maxwell, M. Fatehi Hassanabad, T. Chang, N. Pirani, D. Bona, A.M. Edwards y A.R. Davidson, *Structural and Functional Studies of gpX of Escherichia coli Phage P2 Reveal a Widespread Role for LysM Domains in the Baseplates of Contractile-Tailed Phages*. Journal of Bacteriology, 2013. **195**(24): p. 5461-5468.

84. L.M. Koharudin, A.R. Viscomi, B. Montanini, M.J. Kershaw, N.J. Talbot, S. Ottonello y A.M. Gronenborn, *Structure-function analysis of a CVNH-LysM lectin expressed during plant infection by the rice blast fungus Magnaporthe oryzae*. *Structure*, 2011. **19**(5): p. 662-674.
85. P. Smit, E. Limpens, R. Geurts, E. Fedorova, E. Dolgikh, C. Gough y T. Bisseling, *Medicago LYK3, an entry receptor in rhizobial nodulation factor signaling*. *Plant Physiol*, 2007. **145**(1): p. 183-191.
86. A.A. Nickson, K.E. Stoll y J. Clarke, *Folding of a LysM domain: Entropy-enthalpy compensation in the transition state of an ideal two-state folder*. *Journal of Molecular Biology*, 2008. **380**(3): p. 557-569.
87. B. Miroux y J.E. Walker, *Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels*. *J Mol Biol*, 1996. **260**(3): p. 289-298.
88. F.A. Marston, *The purification of eukaryotic polypeptides synthesized in Escherichia coli*. *Biochem J*, 1986. **240**(1): p. 1-12.
89. A.A. Gust, R. Biswas, H.D. Lenz, T. Rauhut, S. Ranf, B. Kemmerling, F. Götz, E. Glawischnig, J. Lee, G. Felix y T. Nürnberger, *Bacteria-derived Peptidoglycans Constitute Pathogen-associated Molecular Patterns Triggering Innate Immunity in Arabidopsis*. *Journal of Biological Chemistry*, 2007. **282**(44): p. 32338-32348.
90. T. Ohnuma, S. Onaga, K. Murata, T. Taira y E. Katoh, *LysM domains from Pteris ryukyuensis chitinase-A: a stability study and characterization of the chitin-binding site*. *J Biol Chem*, 2008. **283**(8): p. 5178-5187.
91. X.C. Zhang, X. Wu, S. Findley, J. Wan, M. Libault, H.T. Nguyen, S.B. Cannon y G. Stacey, *Molecular evolution of lysin motif-type receptor-like kinases in plants*. *Plant Physiol*, 2007. **144**(2): p. 623-636.
92. E. Gordon, N. Mouz, E. Duee y O. Dideberg, *The crystal structure of the penicillin-binding protein 2x from Streptococcus pneumoniae and its acyl-enzyme form: implication in drug resistance*. *J Mol Biol*, 2000. **299**(2): p. 477-485.
93. A. Dessen, N. Mouz, E. Gordon, J. Hopkins y O. Dideberg, *Crystal structure of PBP2x from a highly penicillin-resistant Streptococcus pneumoniae clinical isolate: a mosaic framework containing 83 mutations*. *J Biol Chem*, 2001. **276**(48): p. 45106-45112.
94. J.V. Olsen, B. Blagoev, F. Gnab, B. Macek, C. Kumar, P. Mortensen y M. Mann, *Global, in vivo, and site-specific phosphorylation dynamics in signaling networks*. *Cell*, 2006. **127**(3): p. 635-648.
95. J. Ciesla, T. Fraczyk y W. Rode, *Phosphorylation of basic amino acid residues in proteins: important but easily missed*. *Acta Biochim Pol*, 2011. **58**(2): p. 137-148.
96. J. Deutscher y M.H. Saier, Jr., *Ser/Thr/Tyr protein phosphorylation in bacteria - for long time neglected, now well established*. *J Mol Microbiol Biotechnol*, 2005. **9**(3-4): p. 125-131.
97. E. Madec, A. Laszkiewicz, A. Iwanicki, M. Obuchowski y S. Seror, *Characterization of a membrane-linked Ser/Thr protein kinase in Bacillus subtilis, implicated in developmental processes*. *Mol Microbiol*, 2002. **46**(2): p. 571-586.
98. M. Fiuza, M.J. Canova, I. Zanella-Cleon, M. Becchi, A.J. Cozzone, L.M. Mateos, L. Kremer, J.A. Gil y V. Molle, *From the characterization of the four serine/threonine protein kinases (PknA/B/G/L) of Corynebacterium glutamicum toward the role of PknA and PknB in cell division*. *J Biol Chem*, 2008. **283**(26): p. 18099-18112.
99. C. Absalon, M. Obuchowski, E. Madec, D. Delattre, I.B. Holland y S.J. Seror, *CpgA, EF-Tu and the stressosome protein YezB are substrates of the Ser/Thr kinase/phosphatase couple, PrkC/PrpC, in Bacillus subtilis*. *Microbiology*, 2009. **155**(Pt 3): p. 932-943.
100. E. Madec, A. Stensballe, S. Kjellstrom, L. Cladiere, M. Obuchowski, O.N. Jensen y S.J. Seror, *Mass spectrometry and site-directed mutagenesis identify several*

- autophosphorylated residues required for the activity of PrkC, a Ser/Thr kinase from Bacillus subtilis*. J Mol Biol, 2003. **330**(3): p. 459-472.
101. A. Ruggiero, B. Tizzano, E. Pedone, C. Pedone, M. Wilmanns y R. Berisio, *Crystal structure of the resuscitation-promoting factor (DeltaDUF)RpfB from M. tuberculosis*. J Mol Biol, 2009. **385**(1): p. 153-162.
 102. A. Ruggiero, D. Marasco, F. Squeglia, S. Soldini, E. Pedone, C. Pedone y R. Berisio, *Structure and functional regulation of RipA, a mycobacterial enzyme essential for daughter cell separation*. Structure, 2010. **18**(9): p. 1184-1190.
 103. W.A. Hoogerwerf, M. Sinha, A. Conesa, B.A. Luxon, V.B. Shahinian, G. Cornelissen, F. Halberg, J. Bostwick, J. Timm y V.M. Cassone, *Transcriptional profiling of mRNA expression in the mouse distal colon*. Gastroenterology, 2008. **135**(6): p. 2019-2029.
 104. I.M. Shah y J. Dworkin, *Induction and regulation of a secreted peptidoglycan hydrolase by a membrane Ser/Thr kinase that detects muropeptides*. Mol Microbiol, 2010. **75**(5): p. 1232-1243.
 105. J. Dworkin y I.M. Shah, *Exit from dormancy in microbial organisms*. Nat Rev Microbiol, 2010. **8**(12): p. 890-896.
 106. A.P. Hendrickx, J.M. Budzik, S.Y. Oh y O. Schneewind, *Architects at the bacterial surface - sortases and the assembly of pili with isopeptide bonds*. Nat Rev Microbiol, 2011. **9**(3): p. 166-176.
 107. R.J. Doyle, J. Chaloupka y V. Vinter, *Turnover of cell walls in microorganisms*. Microbiol Rev, 1988. **52**(4): p. 554-567.
 108. J. Mauck, L. Chan y L. Glaser, *Turnover of the cell wall of Gram-positive bacteria*. J Biol Chem, 1971. **246**(6): p. 1820-1827.
 109. I.M. Shah, M.H. Laaberki, D.L. Popham y J. Dworkin, *A eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments*. Cell, 2008. **135**(3): p. 486-496.
 110. F. Squeglia, R. Marchetti, A. Ruggiero, R. Lanzetta, D. Marasco, J. Dworkin, M. Petoukhov, A. Molinaro, R. Berisio y A. Silipo, *Chemical basis of peptidoglycan discrimination by PrkC, a key kinase involved in bacterial resuscitation from dormancy*. J Am Chem Soc, 2011. **133**(51): p. 20676-20679.
 111. P. Paracuellos, A. Ballandras, X. Robert, R. Kahn, M. Herve, D. Mengin-Lecreulx, A.J. Cozzone, B. Duclos y P. Gouet, *The extended conformation of the 2.9-A crystal structure of the three-PASTA domain of a Ser/Thr kinase from the human pathogen Staphylococcus aureus*. J Mol Biol, 2010. **404**(5): p. 847-858.
 112. A. Ruggiero, F. Squeglia, D. Marasco, R. Marchetti, A. Molinaro y R. Berisio, *X-ray structural studies of the entire extracellular region of the serine/threonine kinase PrkC from Staphylococcus aureus*. Biochem J, 2011. **435**(1): p. 33-41.
 113. W. Hermann, B. Eggers, H. Barthel, D. Clark, T. Villmann, S. Hesse, F. Grahmann, H.J. Kuhn, O. Sabri y A. Wagner, *Correlation between automated writing movements and striatal dopaminergic innervation in patients with Wilson's disease*. J Neurol, 2002. **249**(8): p. 1082-1087.
 114. P. Guntert, *Automated NMR structure calculation with CYANA*. Methods Mol Biol, 2004. **278**: p. 353-378.
 115. R. Koradi, M. Billeter y K. Wuthrich, *MOLMOL: a program for display and analysis of macromolecular structures*. J Mol Graph, 1996. **14**(1): p. 51-55, 29-32.
 116. K. Arnold, L. Bordoli, J. Kopp y T. Schwede, *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling*. Bioinformatics, 2006. **22**(2): p. 195-201.
 117. L. Calvanese, L. Falcigno, C. Maglione, D. Marasco, A. Ruggiero, F. Squeglia, R. Berisio y G. D'Auria, *Structural and binding properties of the PASTA domain of PonA2, a key penicillin binding protein from Mycobacterium tuberculosis*. Biopolymers, 2014. **101**(7): p. 712-719.

CAPÍTULO 4

**CARACTERIZACIÓN DEL
RECONOCIMIENTO DIFERENCIAL
DE LOS ANÓMEROS DE
GALACTOSA POR LECTINAS. EL
CASO DE LA VISCUMINA**

Capítulo 4:

Caracterización del reconocimiento diferencial de los anómeros de galactosa por lectinas. El caso de la viscumina

1 INTRODUCCIÓN Y ANTECEDENTES

La importancia de la presencia de ambas formas anoméricas en medios biológicos se pone de manifiesto por la existencia de enzimas mutarrotasas [1]. También se ha observado selectividad distinta hacia los anómeros en diversos procesos enzimáticos implicados en el metabolismo de carbohidratos [2-3]. Sin embargo, la caracterización estructural del reconocimiento diferencial de anómeros por lectinas o receptores de carbohidratos en solución ha sido poco estudiada. Este hecho tiene que ver con que las

lectinas se unen a los carbohidratos sin modificarlos, y por tanto, no generan respuestas dependientes del tiempo que puedan ser monitorizadas. Teniendo en cuenta que los anómeros en solución se equilibran en unas pocas horas, la mayoría de las técnicas biofísicas disponibles únicamente permiten obtener valores promedios de los parámetros macroscópicos medibles en los procesos de reconocimiento molecular que implican mezclas de anómeros. Cuando se trata de estudiar independientemente un anómero específico, habitualmente se ha recurrido a la estrategia de modificar químicamente la posición anomérica. Por ejemplo, mediante una O-metilación específica se genera la sustitución más simple posible. Así, se rompe el intercambio anomérico y se generan dos compuestos distintos con la configuración anomérica, α o β , fijada.

La aplicación de métodos de RMN puede permitir la discriminación directa del proceso de reconocimiento de cada anómero en la mezcla. Así, por ejemplo, se ha aplicado la resonancia de ^{13}C para demostrar la preferencia por el anómero β de las proteínas de unión a azúcares implicadas en quimotaxis bacteriana [4]. Estudios, más exóticos, de resonancia de tritio (^3H -NMR) se han usado para determinar la preferencia anomérica de la proteína de unión a maltosa [5]. Recientemente, en nuestro grupo de trabajo, se ha demostrado que también la resonancia de ^{19}F se puede aplicar para determinar la selectividad anomérica frente a un receptor [6].

1.1 La viscumina: una lectina vegetal que reconoce galactósidos

En el presente trabajo, se ha elegido la lectina del muérdago europeo (*Mistletoe Lectin* ML-I, también denominada viscumina) como modelo para profundizar en el conocimiento sobre el modo de unión de esta lectina con sus ligandos. La viscumina presenta dos sitios de unión a carbohidratos y hemos abordado el problema del reconocimiento diferencial de los anómeros α y β de la galactosa (Gal) en los dos sitios de unión mediante métodos de RMN. La RMN permite la observación de señales independientes para cada anómero ya que, aunque están en equilibrio químico, la cinética de intercambio es suficientemente lenta y se pueden observar las señales diferenciadas.

La ML-I es una proteína tóxica, presente en el látex de las hojas de la planta del muérdago *Viscumin album* [7]. Tiene un peso molecular aproximado de 60 kDa y

muestra un alto grado de homología, tanto a nivel estructural como de actividad biológica, con la ricina y con otras lectinas del tipo RIP II, ya comentadas en el capítulo de generalidades [8].

La ML-I es un heterodímero de 2 subunidades, A y B, unidas por puente de disulfuro [9-11] (figura 4.1). La actividad como lectina y la especificidad del reconocimiento molecular está localizada en la subunidad B, de 34 kDa. Esta subunidad reconoce preferentemente la galactosa terminal de los glicolípidos y glicoproteínas de las membranas celulares. La cadena B se pliega en dos dominios homólogos con estructura de trébol- β (“ β -trefoil”), y cada uno contiene cuatro subdominios (λ , α , β y γ). Los dos sitios de unión a galactosa están presentes en los subdominios 1α y 2γ [11-12].

Se han descrito, mediante cristalografía de Rayos X, algunas estructuras 3D de la viscumina, en algunos casos en presencia de ligandos (con galactosa PDB=1PUM; con lactosa PDB=1PUU). En el cristal, los dos monómeros AB muestran un doble pliegue simétrico caracterizado por la presencia de un eje de simetría en el dominio 1 de la cadena B adyacente. De este modo, se establecen contactos polares y no polares entre los subdominios α - γ' , β - β' y γ - α' . Esta organización cuaternaria puede explicar las uniones muy fuertes observadas para ciertos glicodendrimeros cuando interaccionan con la viscumina [13-14] y su alta afinidad frente a ciertas glicoproteínas y células [15-16]. La altísima afinidad observada, reflejada en la necesidad de aplicar una fuerza muy alta para separar la viscumina de sus ligandos, ha sido medida mediante microscopia de fuerzas atómicas [17].

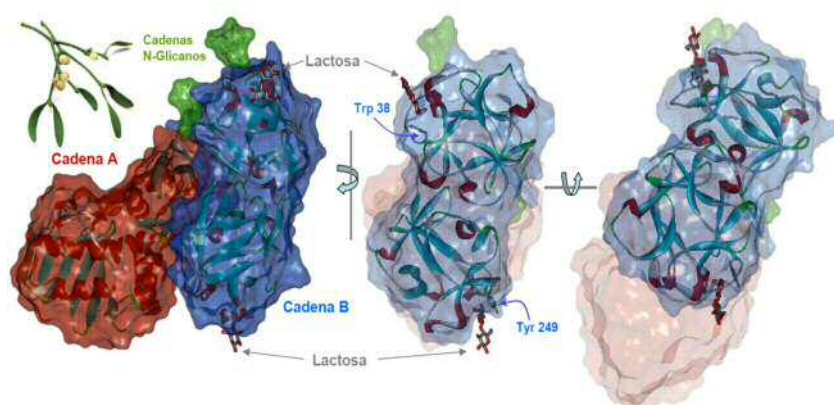


Figura 4.1 Estructura de la viscumina ML-1(PDB 1PUU) extraída del muérdago; en rojo, la cadena A, en verde, se muestran las cadenas de N-glicano; en azul, la cadena B con dos dominios lectina “ β -trefoil”.

En la vista frontal se señalan los 2 sitios de unión a la lactosa en la cadena B: en el primer lóbulo, α_1 =W38, del primer dominio “ β -trefoil” y en el tercer lóbulo, γ_2 =Y249, del segundo dominio “ β -trefoil”.

Después de la unión a los receptores correspondientes, la viscumina se transporta en vesículas endosómicas, dentro de la célula, hacia el aparato de Golgi, y después, hacia el retículo endoplasmático. Aparentemente, en el retículo endoplasmático se produce la ruptura del enlace disulfuro entre las dos subunidades de la Viscumina [18], liberándose así la cadena B de la A. Esta se filtra al interior celular y, mediante su actividad N-glicosidasa sobre los RNA ribosómicos, inhibe el funcionamiento de los ribosomas de forma irreversible, lo que conduce a la muerte celular [19].

A pesar de su toxicidad, la ML-I ha abierto importantes expectativas para su aplicación en el tratamiento del cáncer [20-21] gracias a su efecto inmuno-modulador, antimetastático y antitumoral, descubierto en estudios sobre animales y seres humanos [22-23]. No obstante, aunque en distintos países europeos [23-25] se pueden encontrar distintos extractos del muérdago comerciales que se aplican para el tratamiento del cáncer humano (Iscador, Helixor, Isorel, Abnoba, Icucin,..., etc), su eficacia real es muy controvertida [26].

- La arquitectura del sitio de unión α_1 (Figura 4.2)

Las estructuras cristalinas descritas para los complejos de ML-I con galactosa (1PUM) y lactosa (1PUU) han permitido caracterizar los elementos estructurales para la interacción carbohidrato-proteína [27]. El sitio de unión a Gal en la cadena B de ML-I está formado por Asp23B, por un bucle retorcido Asp23-Val24B-Arg25B y por Gln36B y Lys41B. Cuatro residuos están implicados en la unión a la Gal a través de enlaces de hidrógeno. En el centro del sitio de unión, la cadena lateral de Asp23B se coloca adecuadamente gracias a un enlace de hidrógeno con Gln36B NE2. Esta orientación de la Asp23B asegura el establecimiento de enlaces de hidrógeno entre el Gal 4-OH y Asp23B OD1 así como entre el hidroxilo 3-OH con Asp23B OD2. Además, Gal 4-OH forma enlaces de hidrógeno con Gln36B NE2 y con 26B N. Además, en el entorno de los grupos Gal 2-OH y 3-OH está la Lys41B NZ, que forma enlaces de hidrogeno con ambos. Esta última representa una interacción raramente observada con Gal O2 [27-28].

El anillo aromático de Trp38B forma la parte superior del sitio de unión al azúcar, y proporciona una interacción de apilamiento con el anillo de galactopiranososa. Este

motivo estructural, de interacción de apilamiento CH- π , se observa frecuentemente en los sitios de reconocimiento de lectinas [29].

La orientación de la cadena lateral de triptófano se estabiliza adicionalmente mediante un enlace de hidrógeno a una molécula de agua. Los planos del anillo de triptófano de la proteína y de la piranosa del ligando forman un ángulo de aproximadamente 20° en el caso del complejo ML-I/galactosa.

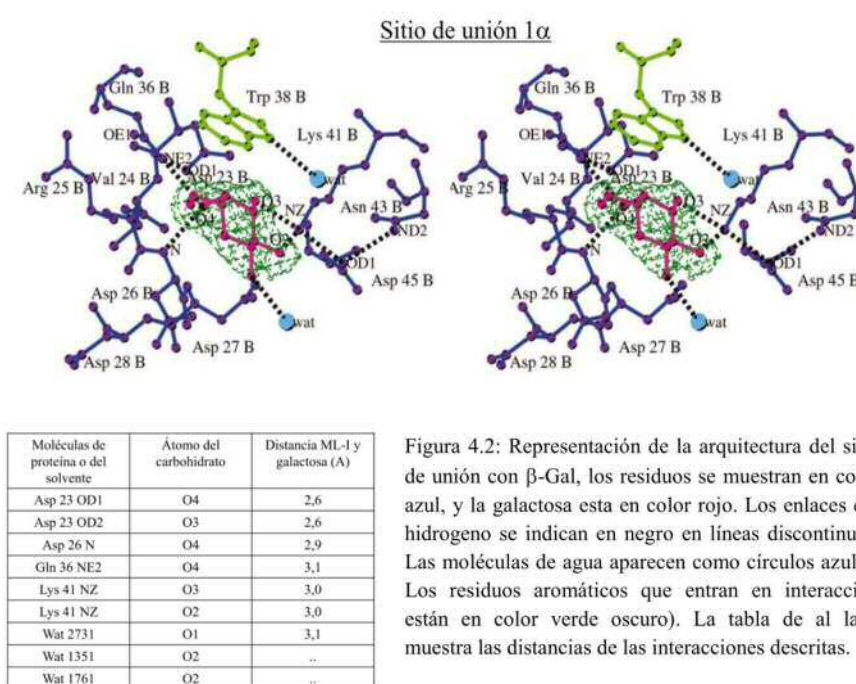


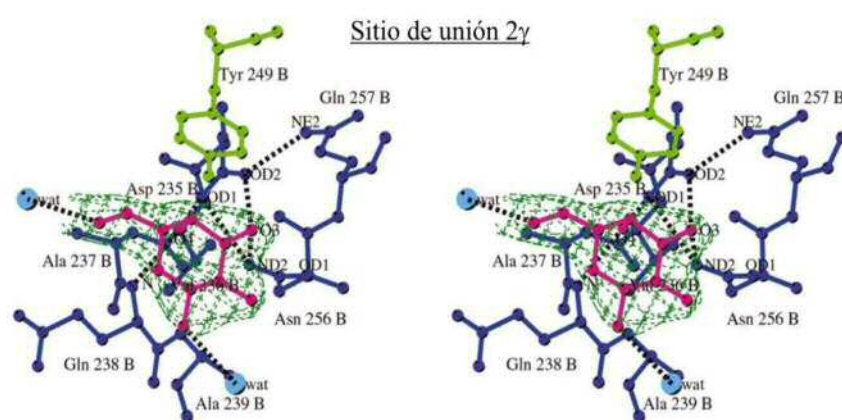
Figura 4.2: Representación de la arquitectura del sitio de unión con β -Gal, los residuos se muestran en color azul, y la galactosa esta en color rojo. Los enlaces del hidrogeno se indican en negro en líneas discontinuas. Las moléculas de agua aparecen como círculos azules. Los residuos aromáticos que entran en interacción están en color verde oscuro). La tabla de al lado muestra las distancias de las interacciones descritas.

- La arquitectura del sitio de unión $\gamma 2$ (Figura 4.3)

La arquitectura general del sitio de unión $\gamma 2$ es casi idéntica al $\alpha 1$, aunque no se conservan todos los aminoácidos implicados en la unión de Gal. El centro del sitio de unión está formado por Asp235B, que se coloca adecuadamente mediante un enlace de hidrógeno entre su OD2 y el NE2 de la Gln257B. El bucle retorcido formado por Asp235B-Val236B-Ala237B se coloca en la parte inferior del sitio de reconocimiento.

La cadena lateral de Asp235B forma dos enlaces de hidrógeno con los hidroxilos 4-OH de Gal (OD1) y 3-OH (OD2). Una mayor estabilización del complejo azúcar-lectina se

consigue mediante enlaces de hidrógeno adicionales entre Gal 4-OH y N Gln238B y entre Gal 3-OH y Asn256B ND2. La parte superior del sitio de interacción está cubierta por Tyr249B, que ofrece la interacción de apilamiento clave con el anillo del azúcar. En este caso, el anillo de galactopiranososa forma un ángulo de aproximadamente 30° con el anillo aromático de Tyr249B. En γ_2 no se observa ninguna interacción entre Gal 2-OH y la lectina [27].



Moléculas de proteína o del solvente	Átomo del carbohidrato	Distancia ML-I y galactosa (Å)
Asp 235 OD1	O4	2,6
Asp 235 OD2	O3	2,7
Gln 238 N	O4	3,0
Asn 256 ND2	O3	3,1
Wat 1571	O6	3,4
Wat 3541	O1	3,0

Figura 4.3: Representación de la arquitectura del sitio de unión con el azúcar, los residuos se muestran en color azul, y la galactosa está en color rojo. Los enlaces del hidrógeno se indican en negro en líneas discontinuas. Las moléculas de agua aparecen como círculos azules. Los residuos aromáticos que participan en la interacción están en color verde). La tabla de al lado muestra las distancias de las interacciones descritas.

La afinidad de la ML-I por galactósidos se ha estudiado mediante diversas técnicas [25, 30-31]. Sin embargo, hay que hacer notar que incluso en el caso más simple del propio monosacárido Gal y de sus α - y β - metil glicósidos (α -metil Gal y β -metil Gal) se han publicado resultados contradictorios. De hecho, Lee y colaboradores, mediante estudios de inhibición, dedujeron que tanto la α -metil Gal como la β -metil Gal inhibían de manera similar la unión de la ML-I a una resina de sepharosa lactosilada [31]. Sin embargo, más tarde Gabius y Surolia encontraron, mediante estudios de calorimetría, que el α -metil Gal tenía una afinidad a ML-I un orden de magnitud superior a la del β -

metil Gal y a la de la propia Gal (15000 M^{-1} para el α -metil Gal frente a 1080 M^{-1} para el β -metil Gal y 1020 para la Gal) [30]). Con estos antecedentes, se eligió la viscumina para estudiar, mediante RMN y de manera individualizada, la diferente selectividad de una lectina hacia cada uno de los dos anómeros α - y β -, que están presentes simultáneamente y en equilibrio químico en las disoluciones del monosacrido.

1.2 Hacia la diferenciación de anómeros: los equilibrios de unión frente a un receptor en el caso de sustratos mutuamente competitivos que están en equilibrio

Antes de iniciar un estudio individualizado de las afinidades de la viscumina por cada uno de los anómeros, es necesario realizar un análisis de los equilibrios presentes en el sistema. La mezcla anomérica en equilibrio se puede considerar como un sistema en el que existe una mezcla de sustratos en equilibrio y, por tanto, que muestra una relación constante de concentraciones durante el proceso de medida. En este caso, las curvas de valoración individualizadas para cada uno de los sustratos no pueden tratarse como en el caso de un modelo sencillo sustrato-receptor, ya que el segundo sustrato actúa como un inhibidor competitivo de la unión del primer sustrato. Es decir, cada uno de los anómeros compete con el otro por la unión a la lectina. Igualmente, las isotermas de Langmuir no son directamente analizables como en el caso de una estequiometría sustrato-receptor 1:1 en presencia de un inhibidor competitivo a concentración constante. En el caso del equilibrio anomérico, la concentración del anómero que actúa como inhibidor varía, de manera asociada, con la concentración del anómero observado. Sin embargo, en el caso sencillo de un receptor (R) en presencia de dos ligandos alternativos (A y B), que mantienen una relación constante ($A/B = \text{cte}$) se puede derivar una relación sencilla entre sus constantes de afinidad y la razón de las concentraciones, e, si se dan las condiciones experimentales en las que la concentración de cada uno de los anómeros está en gran exceso respecto a la concentración del receptor:



$$[At] / [Bt] = c$$

$$\text{Si } [At], [Bt] \gg [Rt] \text{ entonces } [A] \approx [At] \text{ y } [B] \approx [Bt] \text{ y } [A]/[B] = c$$

$$\text{Por tanto, se puede derivar la relación: } [RA] / [RB] = c \times k_{dB} / k_{dA}$$

Por tanto, si se conoce la relación entre las formas unidas de cada uno de los anómeros, y se conoce el equilibrio anomérico, es posible calcular la relación entre las afinidades de ambos anómeros.

Con este objetivo, en este trabajo se ha analizado, mediante técnicas de RMN y modelado molecular, la interacción de la cadena B de la lectina del muérdago ML-I con los ligandos galactosa en sus configuraciones α y β . Primero, y a partir de datos experimentales de titulación obtenidos mediante STD-RMN, se han calculado las constantes de disociación de los ligandos. Estos valores han sido utilizados posteriormente en experimentos de competición para comprobar si los ligandos compiten entre sí en los sitios de unión de la proteína. A continuación, mediante el uso de dos aproximaciones distintas basadas en el cálculo de las velocidades iniciales y del “factor q” [32], se ha descrito el epitopo de unión de cada anómero. El análisis se ha basado en las intensidades de saturación STD transferidas a los protones de los ligandos por su proximidad a la proteína, apoyadas en métodos de modelado molecular (mecánica molecular y “docking”) y en uso del programa CORCEMA (**C**OMplete **R**elaxation and **C**onformational **E**xchange **M**atrix **A**nalysis) [33-35].

2 RESULTADOS

El análisis de la unión de la galactosa (Gal) en solución y de sus correspondientes formas metiladas en la posición anomérica (α y β -metil-Gal) a la viscumina se ha analizado aplicando experimentos de diferencia de transferencia de saturación (STD-RMN). La Gal muestra un equilibrio anomérico $\beta:\alpha$ con una proporción 66:34.

En primer lugar se realizaron experimentos de STD, como control, usando los ligandos libres en ausencia de la viscumina. Se comprobó que los protones de los ligandos no recibían saturación indirecta cuando se irradiaba en la zona alifática (δ 0,85 ppm) del espectro en ausencia de proteína. Se eligió esta frecuencia ya que aquí no aparecen

señales de protones del ligando, por lo que se puede usar para irradiar selectivamente señales de RMN de la viscumina en los experimentos en presencia de la misma.

2.1 Determinación de la constante de disociación de los ligandos con la VAA mediante experimentos STD

Para cada protón cuya intensidad pueda ser medida de manera diferenciada, se obtiene un valor correspondiente de STD. En el caso de que los desplazamientos químicos de distintos protones solapen, sólo se pueden obtener valores promediados de saturación.

En el caso de la Gal libre (figura 4.4), los protones que presentan cierto solapamiento en el espectro ^1H -RMN son los H6 (a,b) de ambas formas α y β -Gal y el H5 de la β -Gal, que aparecen en la zona del espectro entre 3.65 y 3.72 ppm. Sin embargo, es posible obtener información cualitativa de las intensidades de saturación correspondiente a cada protón de cada anómero. En el caso de la α -metil-Gal (figura 4.5), H2 y H3 solapan entre sí, mientras que también lo hacen H6(a, b). En el caso de la β -metil-Gal (figura 4.6) únicamente solapan los protones H6 (a, b) y sus valores de STD se promedian.

En las siguientes figuras (4.4, 4.5 y 4.6) se presentan los espectros de STD de distintas mezclas lectina-ligando. La galactosa (concentración 3.6 mM) se encuentra en su estado de equilibrio anomérico entre las configuraciones α y β . Los α y β -metil-galactósidos se presentan concentraciones de 4.4 y 3.4 mM, respectivamente. La viscumina tiene una concentración de 50 μM . En todos los casos, los experimentos STD se realizaron de manera alternada adquiriendo un número pequeño de FIDs que correspondían a la irradiación (normalmente durante 2 s) a δ 50 ppm, lejos de cualquier señal de resonancia tanto del ligando como del receptor (espectro "off-resonance"). Seguidamente, se adquirió el mismo número de FIDs correspondientes a la irradiación selectiva durante el mismo tiempo (normalmente 2 s) a una frecuencia (δ 0.85 ppm) en la que existen exclusivamente señales de la viscumina en la zona alifática del espectro (espectro "on-resonance"). La recogida de FIDs se realizó, de manera alternada entre ambas frecuencias de irradiación, hasta alcanzar una relación señal/ruido satisfactoria del espectro STD.

El espectro STD se obtiene sustrayendo el espectro "on-resonance" (I_{sat}), en el cual se ha saturado la proteína, del espectro "off-resonance" (I_0), obtenido como referencia. La

amplitud de la saturación de cada protón se expresa en forma de porcentaje de la pérdida de señal ($I_0 - I_{\text{sat}}$) respecto al espectro "off-resonance", (I_0).

$$A_{\text{STD}} = \frac{I_0 - I_{\text{sat}}}{I_0} \times 100 = \frac{I_{\text{STD}}}{I_0} \times 100 \quad (1)$$

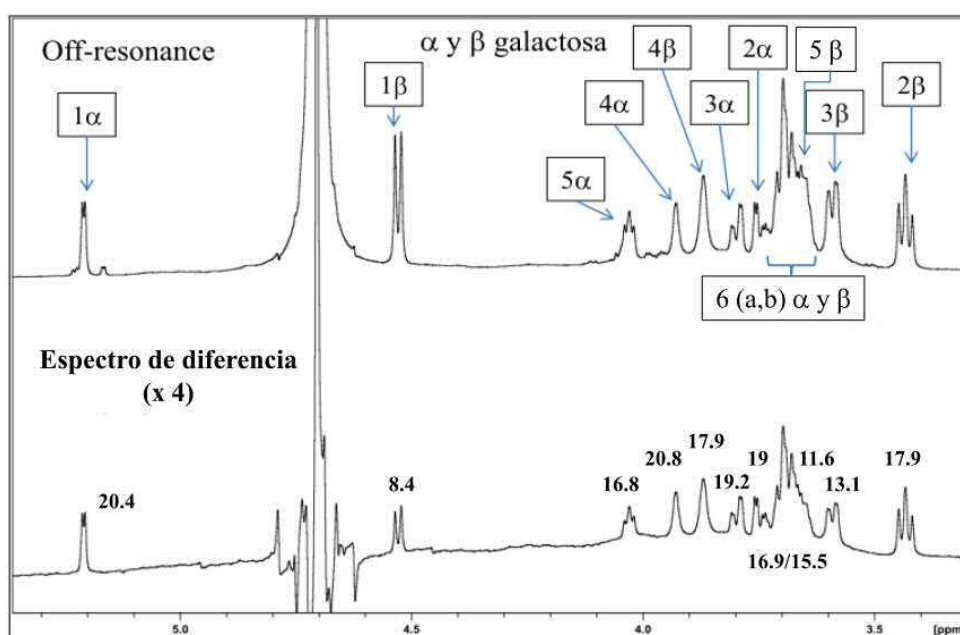


Figura 4.4 Espectro STD (panel inferior) de la Gal (3,6 mM) en presencia de la VAA (0,05 mM) en disolución. El panel superior muestra el espectro de referencia "off-resonance", con la correspondiente asignación de los protones de la mezcla anomérica. La intensidad del espectro STD está ampliada por un factor de 4 y se indican los % de saturación de cada señal.

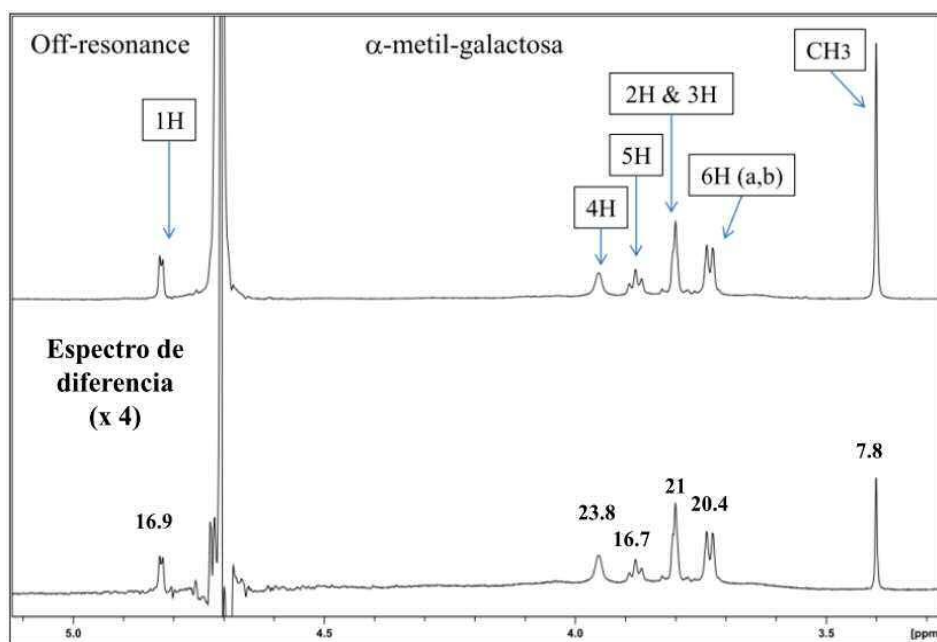


Figura 4.5 Espectro STD (panel inferior) de α metil Gal (4,4 mM) en presencia de VAA (0,05 mM) en disolución. El panel superior muestra el espectro “off-resonance”.

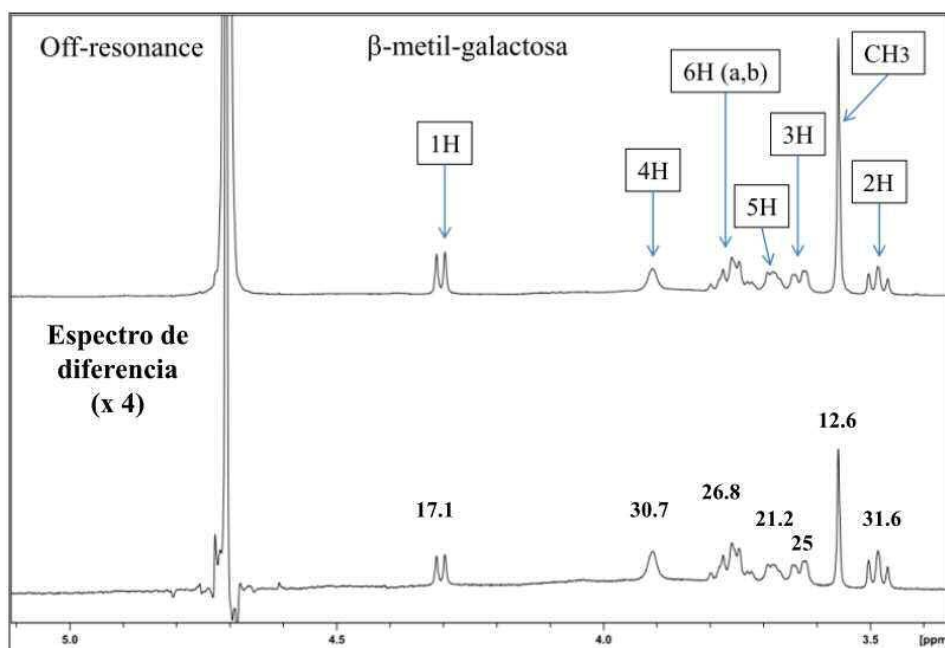


Figura 4.6 Espectro STD (panel inferior) de la β metil Gal (3,4 mM) en presencia de VAA (0,05 mM) en disolución. El panel superior es el espectro “off-resonance”.

Para obtener las constantes de disociación, se siguió una metodología similar a la descrita por Angulo et al. [36]. Esta se basa en la cuantificación de distintos experimentos de STD, obtenidos para distintos tiempos de saturación. Así, se elaboran curvas de crecimiento de saturación para cada protón del ligando. En concreto, se realizaron series de experimentos de titulación usando diversos tiempos de saturación (0.5, 1, 2, 4, 8 s) y diferentes concentraciones de ligando, y empleando una concentración fija de viscumina (50 μM , figura 4.7).

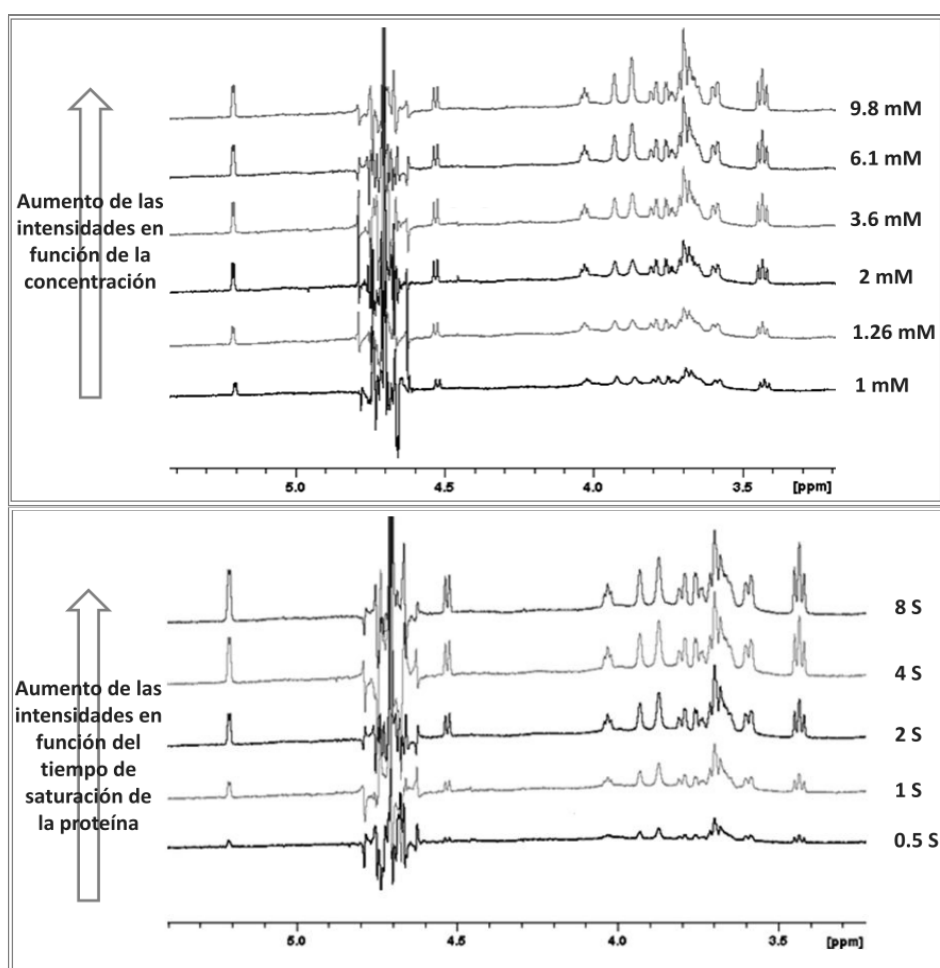


Figura 4.7 En la serie de espectros del panel superior se muestra el crecimiento de las señales de ^1H -STD de Gal para un tiempo fijo de saturación (2s) y concentraciones variables del ligando. En los espectros del panel inferior se muestra el crecimiento de las señales STD usando una concentración fija de Gal (6,1 mM) en presencia de viscumina (50 μM), empleando diferentes tiempos de saturación

A partir de los experimentos de STD a distintos tiempos de saturación y con distintas concentraciones de ligando, se calcularon, en primer lugar, los porcentajes de saturación de cada uno de los protones de los ligandos. Seguidamente se determinó la concentración saturada de cada protón (STD) obtenida para cada tiempo de saturación, multiplicando la saturación relativa (I_{sat}/I_0) por la concentración de ligando correspondiente. Con estos valores se construyeron las curvas de crecimiento de saturación, que se ajustaron a una función de exponencial creciente (2). Como parámetros ajustables se usaron la saturación máxima (STD_{max}) alcanzable para la concentración de lectina usada y una constante aparente de transferencia de saturación (k_{sat}). A partir de la pendiente inicial de estas curvas, se obtuvieron las velocidades iniciales de saturación para cada protón, como el producto de ambos parámetros STD_{max} por k_{sat} . (ver materiales y métodos) [36].

$$\text{STD}(t_{\text{sat}}) = \text{STD}_{\text{max}} \cdot (1 - e^{k_{\text{sat}} \cdot t_{\text{sat}}}) \quad (2)$$

$$V_{i_{\text{sat}}} = \text{STD}_{\text{max}} \times k_{\text{sat}} \quad (3)$$

A modo de ejemplo, los resultados obtenidos para el protón H4 del anómero α -Gal se presentan en la figura 4.8. Los resultados obtenidos para el resto de los protones se muestran en el anexo I (figuras 1, 2, 3 y 4).

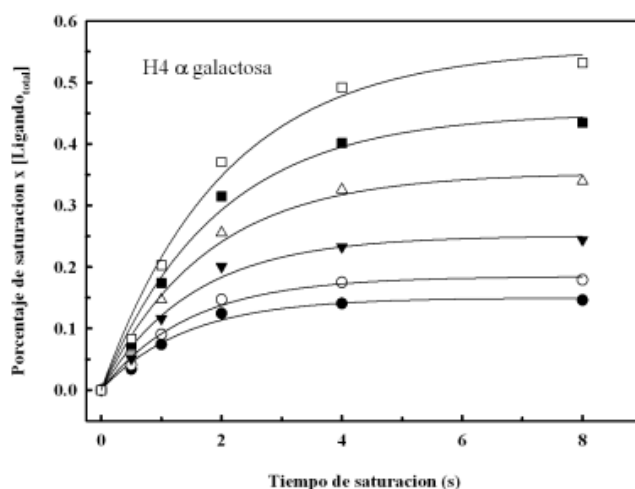


Figura 4.8 Curvas de crecimiento de saturación (expresado como concentración absoluta saturada del protón observado en mM) en función del tiempo de saturación para el protón H4 del anómero α Gal a diferentes concentraciones de Gal (●1; ○ 1,26; ▼ 2; Δ 3,6; ■6,1; □9,8 mM). La concentración del anómero α corresponde a un 33% de la Gal total

A partir de las curvas (figura 4.8) se puede observar que, cuanto mayor es la concentración del ligando, mayor es la concentración absoluta saturada, expresada para cada protón como el producto de la saturación relativa observada (I_{std}/I_0) por la concentración total de ligando; es decir, el valor absoluto de la señal de saturación. Por lo tanto, los experimentos con mayor concentración de ligando presentan mejor sensibilidad (véanse, por ejemplo, los espectros del panel superior de la figura 4.7).

A tiempos de saturación muy cortos, los STD tienen problemas de sensibilidad, tanto por la baja relación señal:ruido, como por la saturación ineficaz de la proteína (figura 4.9). Por otro lado, el uso de tiempos de saturación muy largos asegura la saturación máxima de la lectina, por lo que se obtienen incrementos considerables en las intensidades STD. No obstante, este hecho sólo se da hasta alcanzar un estado estacionario en el que es necesario considerar la relajación longitudinal de la forma no unida [32] y la reasociación del ligando saturado [36] que se acumula en solución [37].

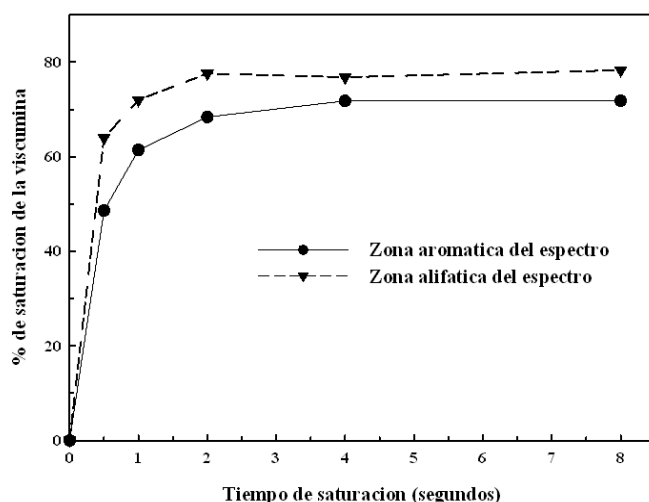


Figura 4.9 Curvas de crecimiento de la saturación de la viscumina en función del tiempo de irradiación, tomando como referencia las intensidades de las señales de STD de la zona aromática y alifática del espectro (irradiación a δ 0,85).

en efecto, según aumenta el tiempo de irradiación, las poblaciones crecientes de los espines saturados del ligando pueden, en principio, volver a unirse al receptor, formando complejos no productivos en términos de transferencia de saturación. Asimismo, en el estado no enlazado, las intensidades empiezan a retornar hacia sus valores de equilibrio mediante procesos de relajación longitudinal. De este modo, la acumulación de saturación para un protón dado en disolución, que, a tiempos cortos de irradiación, se acercaba a una progresión lineal, se ralentiza y muestra una pendiente cada vez menor, hasta alcanzar un estado estacionario. Aquí, la saturación ganada por cada protón se compensa por la relajación que sufre cuando el ligando está en el estado no asociado.

Por otro lado, la observación de las curvas de crecimiento para los distintos protones permite evidenciar (figura 4.10) que existen comportamientos diferenciados, tanto en la saturación máxima alcanzable, como en la velocidad con que se alcanza el estado estacionario a dicha saturación máxima. Este comportamiento diferencial depende, por un lado, de la eficiencia con que la proteína transfiere la saturación a cada protón del ligando y, por otro lado, de la distinta relajación de los protones del ligando, que depende de sus tiempos de relajación longitudinal. Así, por ejemplo, la relajación es

mucho más eficaz para los H6(a,b) de las Gal metiladas, como se puede deducir de la observación de las columnas de tiempos de relajación obtenidos para cada uno de los protones de los ligandos en las tablas 4.5 y 4.6. Este hecho provoca que la saturación observada para estos protones alcance un estado estacionario a tiempos de saturación más cortos que para otros que relajan más lentamente.

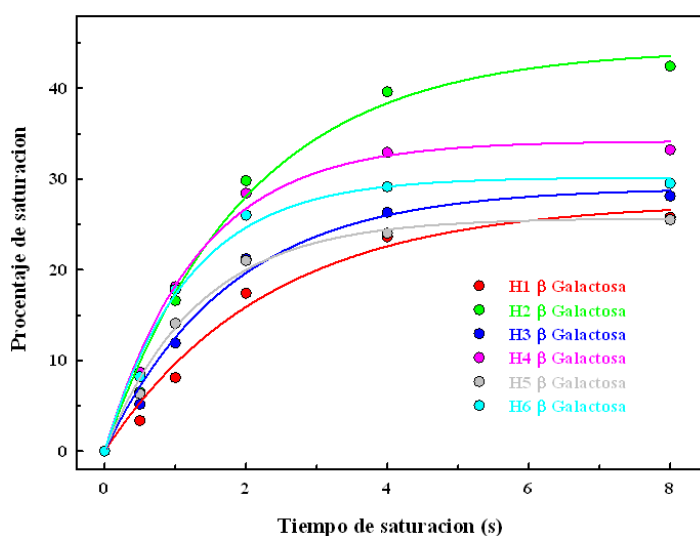


Figura 4.10 Curvas de crecimiento del % de saturación observado para cada protón de la β Gal (0.83 mM) en función del tiempo de irradiación a δ 0.85 ppm. Las curvas continuas corresponden al ajuste de los valores experimentales para cada protón a una función exponencial creciente (ecuación 1). Puede observarse el efecto evidente del tiempo de relajación sobre las intensidades de saturación.

Por tanto, para poder utilizar de una manera cuantitativa las saturaciones de los diversos protones y así evitar el problema de las diferencias de tiempo de relajación de cada protón, se recurrió al cálculo de las velocidades iniciales, según las ecuaciones 2 y 3.

Las velocidades iniciales de saturación para cada protón, en semejanza a un proceso enzimático, son proporcionales a la concentración de forma unida [Proteína-Ligando] y permiten la estimación de las constantes de disociación, mediante su ajuste a ecuaciones hiperbólicas del tipo de isothermas de Langmuir:

$$V_{i_{obs}} = \frac{(V_{i_{max}} * L)}{(L + K_d)} \quad (4)$$

Donde $V_{i_{obs}}$ es la velocidad inicial de saturación observada para cada protón a una concentración L de ligando en la forma no asociada; $V_{i_{max}}$ es la velocidad máxima de saturación a una concentración constante de ligando y K_d es la constante de disociación. Teniendo en cuenta que la concentración de ligando está en exceso sobre la de la proteína, se aplica la aproximación $L=L_t$.

En el caso de la Gal, al estar presente en forma de una mezcla de anómeros en equilibrio, [38] con una relación de concentraciones constante durante el tiempo de medida de la saturación, los datos experimentales de las velocidades iniciales de saturación observados se ajustaron usando la ecuación correspondiente que tiene en consideración el efecto mutuamente competitivo de ambos anómeros entre sí, según se ha descrito anteriormente (apartado 4.1.2).

$$V_{i_{obs}}[RA] = V_{i_{max}} * A / (K_{dA_app} + A) \quad \text{donde } K_{dA_app} = K_{dA}(1+B/K_{dB}) \quad (5)$$

$$V_{i_{obs}}[RB] = V_{i_{max}} * B / (K_{dB_app} + B) \quad \text{donde } K_{dB_app} = K_{dB}(1+A/K_{dA}) \quad (6)$$

$$A/B = \text{constante}$$

Como primera aproximación, se consideró que ambos anómeros tenían la misma afinidad, K_d . En este caso, las ecuaciones individualizadas para cada anómeros se simplifican:

$$V_{i_{obs}}\alpha = \frac{(V_{i_{max}}\alpha * [\alpha])}{([Galactose] + K_d)} \quad \text{y} \quad V_{i_{obs}}\beta = \frac{(V_{i_{max}}\beta * [\beta])}{([Galactose] + K_d)} \quad (7) \text{ y } (8)$$

Sin embargo, al calcular las constantes de disociación con estas ecuaciones, a partir de las velocidades iniciales de saturación para cada protón, se obtuvieron valores distintos de K_d para ambos anómeros (1,4 mM para α -Gal y 1,9 mM para β -Gal).

Adicionalmente, también se consideró la presencia funcional de los dos sitios de unión independientes descritos en la viscumina. Así, las curvas se ajustaron a las correspondientes para un sistema bivalente:

$$Vi_{obs}\alpha = \frac{(Vi_{\max,siteA\alpha} * [\alpha])}{([Galactose] + K_{d,siteA})} + \frac{(Vi_{\max,siteB\alpha} * [\alpha])}{([Galactose] + K_{d,siteB})} \quad (9)$$

$$Vi_{obs}\beta = \frac{(Vi_{\max,siteA\beta} * [\beta])}{([Galactose] + K_{d,siteA})} + \frac{(Vi_{\max,siteB\beta} * [\beta])}{([Galactose] + K_{d,siteB})} \quad (10)$$

Sin embargo, en ninguno de los casos, los ajustes de los datos experimentales a estas ecuaciones bivalentes reprodujeron adecuadamente el comportamiento de las curvas de valoración. Los ajustes, en algunos casos, devolvían los mismos parámetros de V_i y K_d para ambos sitios, o bien asignaban valores sin sentido en términos de equilibrios químicos (constantes de disociación y velocidades exageradamente grandes).

Como segunda aproximación, se consideró que ambos anómeros tenían afinidades distintas, siendo la α -galactosa el anómero con mejor afinidad. Si se asume que ambos anómeros, en lo que se refiere a la posición H4, presentan una misma geometría de contacto con la proteína en el sitio de unión, la relación entre las saturaciones de los protones H4 de ambos anómeros (considerando, además, que los tiempos de relajación T1 son iguales en ambos casos) debería ser proporcional a la relación de las concentraciones unidas de ambos anómeros.

Teniendo en cuenta que la relación entre la fracciones molares de cada uno de los anómeros es inversamente proporcional a la relación de constantes de disociación,

$$\frac{K_{DB}}{K_{DA}} = \frac{[RA]/At}{[RB]/Bt} \frac{\%StdH4\alpha}{\%StdH4\beta} = \frac{[RA]/At}{[RB]/Bt} \quad (11)$$

La premisa anterior se respaldó observando que, para todas las condiciones de concentración y tiempo de saturaciones estudiadas, la relación de las fracciones molares referidas a la saturación de los protones H4 de ambos anómeros era siempre favorable al anómero α (1,15 veces). Obviamente, esta comparación de afinidades entre ambos anómeros no se puede aplicar directamente al resto de los protones, dadas las diferencias en tiempos de relajación y/o distintas geometrías en el sitio de unión. Sin embargo, en todas las condiciones, se encontró que el porcentaje de saturación de un protón dado en el anómero α era siempre superior a su equivalente en el anómero β .

Teniendo en cuenta esta relación entre las constantes de afinidad, las curvas de saturación obtenidas para cada protón, en función de la concentración de cada anómero, se ajustaron a las ecuaciones derivadas para el caso de la competición mutua entre ambos anómeros, considerando diferentes afinidades para cada uno de ellos.

$$Vi_{obs}\alpha = \frac{(Vi_{max\alpha} * [\alpha])}{(K_{d\alpha} + [\beta]K_{d\alpha} / K_{d\beta} + [\alpha])} \quad Vi_{obs}\beta = \frac{(Vi_{max\beta} * [\beta])}{(K_{d\beta} + [\alpha]K_{d\beta} / K_{d\alpha} + [\beta])} \quad (12) \text{ y } (13)$$

Aún así, fue necesario introducir un término inespecífico, $Ns*[Ligando]$, proporcional a la concentración de ligando, para obtener buenos ajustes matemáticos entre las curvas calculadas y los valores experimentales.

$$Vi_{obs}\alpha = \frac{(Vi_{max\alpha} * [\alpha])}{(K_{d\alpha} + [\beta]K_{d\alpha} / K_{d\beta} + [\alpha])} + Ns * [\alpha] \quad (14)$$

$$Vi_{obs}\beta = \frac{(Vi_{max\beta} * [\beta])}{(K_{d\beta} + [\alpha]K_{d\beta} / K_{d\alpha} + [\beta])} + N_s * [\beta] \quad (15)$$

Los resultados de los ajustes, analizados protón a protón (figuras 4.11 y 4.12), muestran una cierta variabilidad en los valores de las constantes de disociación obtenidas para cada protón (ver tabla 4.1). Por tanto, para comparar ambos compuestos se calculó el valor promedio de las constantes estimadas para cada protón, en cada anómero, obteniendo para la α Gal una $K_d = 0,73 \text{ mM} \pm 0,27$ y para la β Gal, una $K_d = 1,38 \text{ mM} \pm 0,5$.

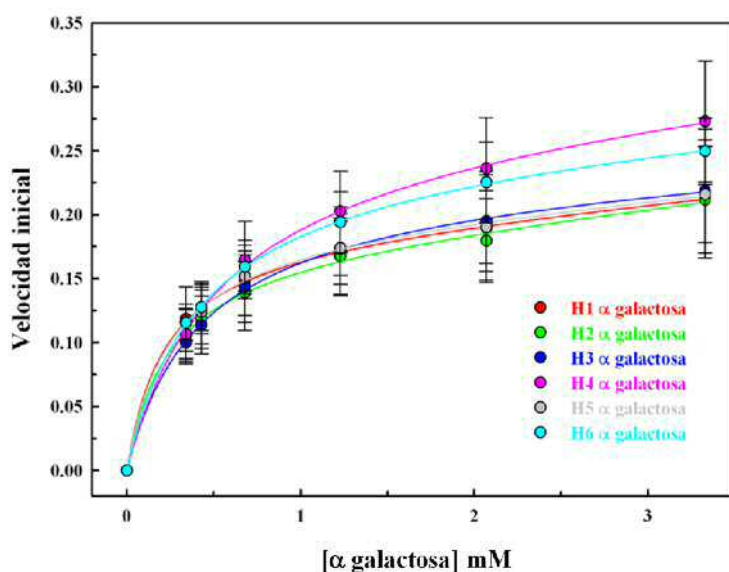


Figura 4.11 Curvas de asociación teórica que relacionan el aumento de la velocidad inicial de interacción de cada protón con la concentración de la α -Gal, según el ajuste a un sitio de unión específico más otro inespecífico (ecuación 14). Las barras de error indican el error calculado en las Vi según la ecuación 2.

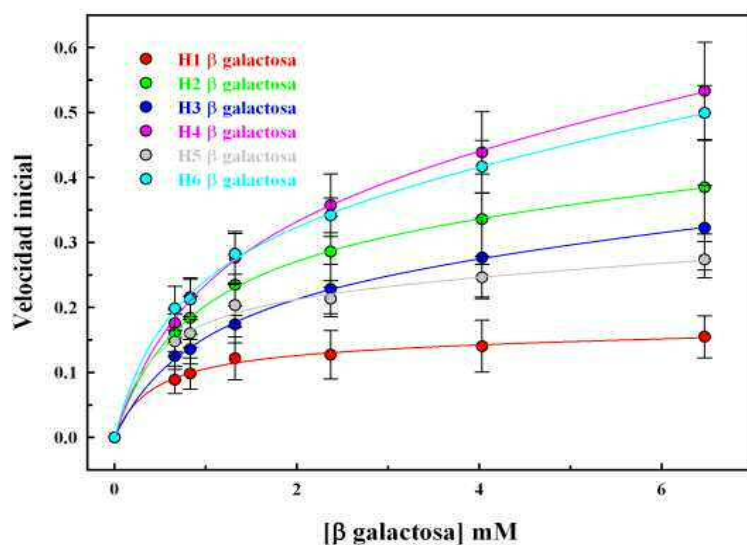


Figura 4.12 Curvas de asociación teórica, que relacionan el aumento de la V_i de interacción de cada protón con la concentración de la β -Gal, según el ajuste a un sitio de unión específico más otro inespecífico (ecuación 14). Las barras de error indican el error calculado en las V_i según la ecuación 2

Tabla 4.1 Valores teóricos de la constante de disociación k_d , obtenida por los protones de la α y β -Gal, según las ecuaciones del ajuste matemático descritas en el texto.

α Gal		H1	H2	H3	H4	H5	H6
1 sitio + factor inespecífico	k_d	0,46	0,50	0,89	1,10	0,51	0,88
	V_{imax}	0,39	0,37	0,47	0,57	0,41	0,52
	Ns	0,012	0,014	0,006	0,013	0,011	0,010
β Gal		H1	H2	H3	H4	H5	H6
1 sitio + factor inespecífico	k_d	0,78	1,58	1,86	1,94	0,81	1,30
	V_{imax}	0,27	0,66	0,53	0,80	0,45	0,69
	Ns	0,002	0,011	0,012	0,025	0,007	0,026

Comparando las saturaciones relativas obtenidas para las distintas parejas de protones de ambos anómeros y el promedio obtenido para las constantes de disociación, se observa cualitativamente y, en términos macroscópicos, una ligera preferencia de la viscumina por el anómero α . Sin embargo, los resultados no permiten diferenciar, a

escala microscópica, ni el porcentaje de participación ni las afinidades parciales de cada uno de los dos sitios de unión de galactósidos de la viscumina.

En un intento adicional por tratar de caracterizar la posible capacidad de discriminación de los anómeros de Gal por la viscumina, se procedió al estudio comparativo, usando STD, de la interacción entre viscumina y los α - y β -metil Gal. Estas son las moléculas más sencillas de galactosa con configuración anomérica fija. En este punto hay que resaltar que Bharadwaj et al. [30] ,a partir de medidas de titulaciones por calorimetría isoterma (ITC) describieron una diferencia de afinidad significativa de la viscumina, favorable al α -metil Gal ($K_d = 0,06$ mM) frente al β -metil Gal (0,93 mM).

De nuevo, se realizaron experimentos de STD con cada uno de los metil derivados por separado, usando varias concentraciones de ligando y distintos tiempos de irradiación. A partir de las curvas de crecimientos de la saturación (figuras 4.13 y 4.14) se calcularon las correspondientes velocidades iniciales de saturación para cada protón, mediante la aproximación a una curva exponencial creciente, como se ha descrito más arriba. Seguidamente, las velocidades iniciales obtenidas se representaron en función de la concentración de ligando y los valores obtenidos se ajustaron a ecuaciones hiperbólicas del tipo de isothermas de Langmuir considerando la existencia en la viscumina de uno:

$$Vi_{obs} = \frac{(Vi_{max} * [L])}{([L] + K_d)}$$

o dos sitios de unión independientes:

$$Vi_{obs} = \frac{(Vi_{max,siteA} * [L])}{([L] + K_{d,siteA})} + \frac{(Vi_{max,siteB} * [L])}{([L] + K_{d,siteB})} \quad (16)$$

Análogamente a lo observado para la Gal, cuando los datos se ajustaron a la ecuación correspondiente a dos sitios, los datos experimentales no discriminaron la posible

existencia de dos sitios de unión con afinidades distintas. En el caso de la α -metil Gal, los resultados los ajustes devolvieron un segundo sitio, pero con unos valores de K_d fuera de rango y sin sentido físico. Los datos experimentales eran insuficientes para caracterizar, de manera fiable, un posible sitio de alta afinidad. Por el contrario, en el caso de la β -metil galactosa, los resultados no indicaron la existencia de un posible sitio de alta afinidad. Los ajustes devolvieron dos sitios equivalentes indistinguibles y con la misma K_d que cuando se usó la ecuación de ajuste a un sitio único.

Finalmente, la introducción de un término inespecífico proporcional a la concentración de ligando permitió mejorar el ajuste matemático de los datos:

$$V_{i_{obs}} = \frac{(V_{i_{max}} * [L])}{([L] + K_d)} + N_s * [L] \quad (17)$$

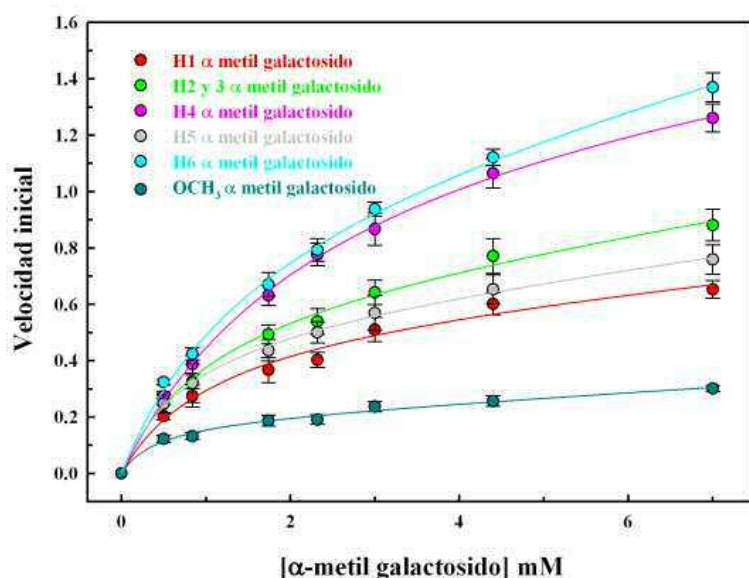


Figura 4.13 Curvas de asociación teórica que relacionan el aumento de la V_i de interacción de cada protón con la concentración de la α -metil Gal según el ajuste a un sitio de unión específico + otro inespecífico (ecuación 17). Las barras de error indican el error calculado en la V_i .

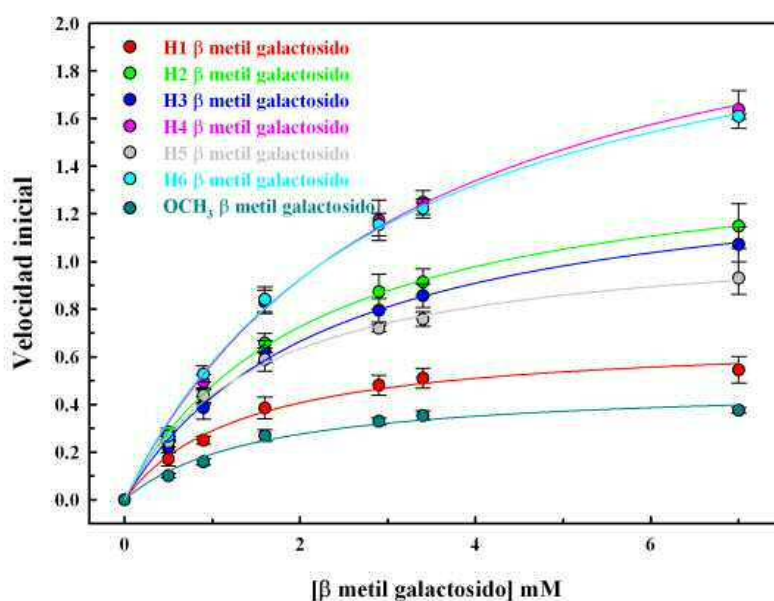


Figura 4.14 Curvas de asociación teórica que relacionan el aumento de V_i de interacción de cada protón con la concentración de la β -metil Gal según el ajuste a un sitio de unión específico + otro inespecífico (ecuación 17). Las barras de error indican el error calculado en la V_i .

La comparación de las curvas de crecimiento ajustadas según las ecuaciones de formación de complejo monovalente, bivalente o bivalente con un sitio inespecífico, permite observar un comportamiento similar entre los ajustes obtenidos para los ligandos α y β -Gal, así como el β -Me-Gal, mientras que, en el caso de la α -Me-Gal, los mejores ajustes de regresión no lineal de los valores de velocidad de transferencia de saturación son distintos. De hecho, corresponden a las ecuaciones estándares que definen la unión bivalente con 2 uniones específicas o bivalente, con una unión específica del ligando al receptor más otra unión inespecífica. Estas se describen por unas hipérbolas rectangulares de color verde y rojo que atraviesan la mayor parte de los puntos.

Tabla 4.2 Valores calculados de la constante de disociación K_d obtenida por los protones de la α - y β -Metil Gal, según el ajuste a la ecuación 17

α -Me-Gal/ML-I		H1	H2 y H3		H4	H5	H6 (ayb)	OCH ₃
1 sitio + factor inespecífico	k_d	1,09	1,03		2,49	1,00	1,59	0,43
	V_{max}	0,56	0,67		1,46	0,56	1,08	0,19
	Ns	0,027	0,044		0,027	0,037	0,07	0,017
β -Me-Gal/ML-I		H1	H2	H3	H4	H5	H6 (ayb)	OCH ₃
1 sitio + factor inespecífico	k_d	1,35	2,13	2,34	3,25	1,42	3,00	1,52
	V_{max}	0,68	1,5	1,44	2,43	1,05	2,31	0,48
	Ns	$1,2e^{-9}$	$1,3e^{-9}$	$7,8e^{-9}$	$8,1e^{-9}$	0,007	$7,5e^{-9}$	$1,04e^{-9}$

Por otro lado, los valores de constante de disociación dieron ventaja a los resultados registrados en el caso de la ecuación bivalente, que corresponde a un sitio específico y otro inespecífico. De este modo, podemos resumir los parámetros de afinidad para α y β Gal, así como para α y β metil Gal en la tabla 4.3.

Los resultados obtenidos, además de dar información sobre la afinidad, en el caso de los valores de V_{max} indican la eficiencia de la transferencia de saturación desde la proteína a los diferentes protones del ligando. Por tanto, están directamente relacionados con la cercanía entre los protones de la proteína y los del ligando, permitiendo extraer información de las zonas del ligando en contacto con la proteína. En la figura 4.15 se presentan los perfiles, para cada uno de los 4 galactósidos, de las velocidades iniciales (normalizadas con respecto a la máxima velocidad) de transferencia de saturación para cada protón del ligando. Se observa que, en todos los casos, los protones H4, seguidos de los protones H6a/b, son los que reciben más eficientemente la saturación.

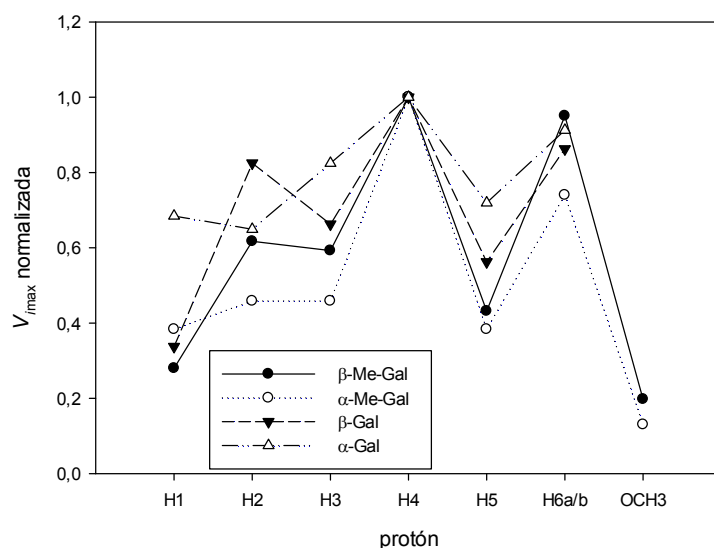


Figura 4.15 Representación de los perfiles de velocidad de saturación normalizados para cada protón de los galactósidos.

Por otro lado, de las tablas 4.1 y 4.2 se seleccionaron los valores obtenidos, en términos de constante de disociación obtenidos, en el caso de la ecuación bivalente que corresponde a un sitio específico y otro inespecífico (ecuación 17), y se extrajo un valor promedio para cada ligando a partir de los valores de cada protón del ligando. De este modo, podemos resumir los parámetros de afinidad para α y β Gal así como para α y β metil Gal en la Tabla 4.3.

Tabla 4.3 Valores promedio de todos los protones de cada ligando a partir de los ajustes que corresponde a una unión específica y otra inespecífica (ecuaciones 14, 15 y 17) del ligando con la proteína

Ligandos	Parámetros calculados		Descritos [30]
	k_d	N_s	k_d
α -Gal	$0,72 \pm 0,3$	0,011	0,98
β -Gal	$1,37 \pm 0,5$	0,012	
α -Me-Gal	$1,27 \pm 0,7$	0,032	0,06
β -Me-Gal	$2,14 \pm 0,8$	$1,47 \text{ e}^{-3}$	0,93

Basados en nuestros cálculos, las afinidades de α y β -Gal hacia la viscumina son de 0,72 y 1,37 mM, respectivamente, y de 1,3 y 2,1 mM para los correspondientes α y β metil-Gal, respectivamente. No obstante, los valores de constantes de disociación obtenidas para cada uno de los protones de un mismo ligando muestran una cierta dispersión. En el caso de la Gal libre, la media de los k_d obtenidas para los dos conformeros es comparable con el obtenido del cálculo de la afinidad descrita mediante experimentos de titulación por calorimetría, en los que no se diferenciaron ambos anómeros (0,98 mM) [30]. Sin embargo, el valor obtenido para la constante de disociación del ligando α -Me-Gal en nuestras condiciones experimentales es más de un orden de magnitud superior al descrito por calorimetría para el mismo ligando (referencia[30]). Por el contrario, en el caso de la β -Me-Gal, aunque el valor calculado en nuestras condiciones experimentales difiere con el descrito, ambos se mantienen en un rango milimolar, con una relación 1:2.

Con el fin de verificar las diferencias de los resultados anteriores con los ya descritos, se decidió analizar la unión de la α y β metil-Gal mediante experimentos de competición.

2.2 Experimentos de competición entre α y β metil galactósidos

Primero se realizó un experimento sencillo para poder comparar, de forma cualitativa, las afinidades de los dos ligandos, α y β metil Gal, cuando se enfrentan a la vez a la viscumina a la misma concentración (1 mM) y usando el mismo tiempo de saturación (2 s). En este experimento, se siguió el comportamiento de los protones H4, asumiendo que en ambos isómeros, los H4 tienen un modo de unión similar y deben presentar una misma geometría de contactos con la proteína en el sitio de unión, según se deduce de los perfiles de velocidades de saturación mostrados en la figura 4.15. Por tanto, la relación entre las saturaciones de los protones H4 de ambos isómeros se puede considerar proporcional a sus afinidades, en mezclas equimoleculares (figura 4.16).

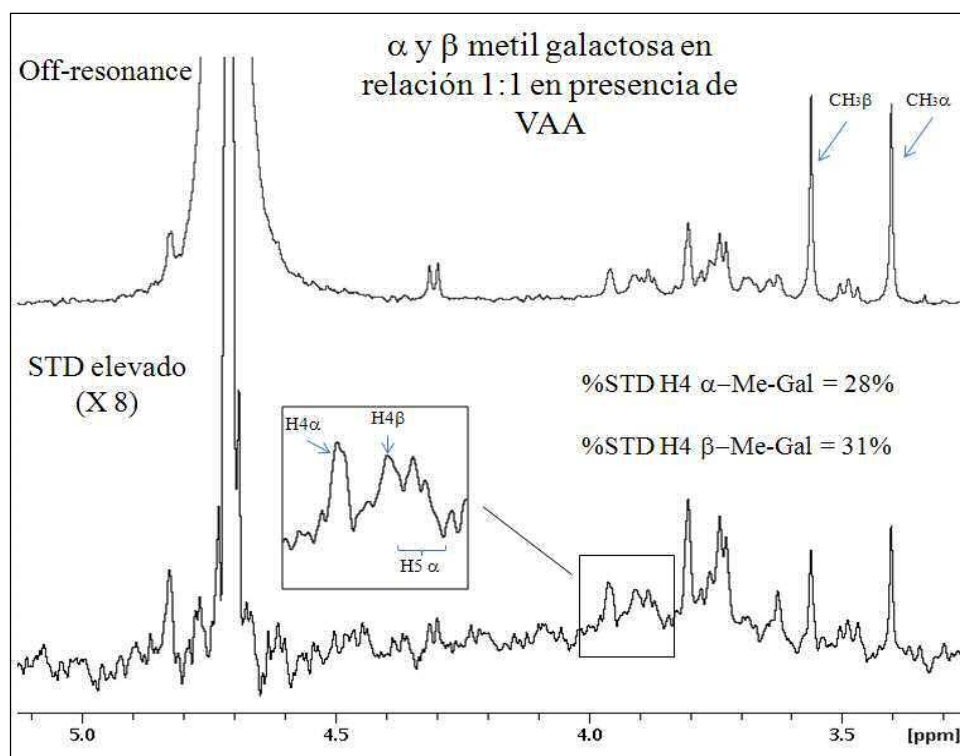


Figura 4.16 Espectro de STD de una mezcla equimolecular de α y β metil Gal (1mM) en presencia de la viscumina. En la figura se indican los porcentajes de saturación del protón H4 de ambos isómeros.

Afortunadamente, no superposición de las señales de α -Me-Gal y β -Me-Gal en el espectro de la mezcla y los protones H4 de ambos isómeros son fácilmente diferenciables. En el espectro STD de esta mezcla se observa una tendencia similar en la saturación de los protones H4. Este resultado es contrario al esperado para la superior afinidad descrita para el ligando alfa-metil Gal ($k_d = 0,06$ mM) que para el β metil-galactosido ($k_d = 0,93$ mM) [30]. Nuestros resultados indican que las dos moléculas tienen afinidades similares por la viscumina, en el rango milimolar de concentraciones, y en nuestras condiciones experimentales.

Seguidamente, con objeto de cuantificar la relación de afinidades de ambos isómeros, se realizó una serie de experimentos de STD añadiendo concentraciones crecientes de β -metil Gal (0, 0.3, 0.6, 1.2, 2.1, 3.9, 8.2 mM) a una muestra con una concentración fija de α -metil Gal (1 mM), en presencia de viscumina. La disminución de la intensidad de la señal de STD de los protones de la molécula observada (α -Me-Gal) en función del

aumento de la concentración del competidor (β -Me-Gal) se ajustó mediante la correspondiente ecuación que se refiere a una inhibición competitiva, permitiendo así el cálculo de las constantes de inhibición para cada protón.

En el caso de este experimento de competición, la constante aparente de disociación de la viscumina, $k'_{d,L}$, para el ligando observado **L** es mayor que la constante de disociación real, $k_{d,L}$, a causa de la unión competitiva del inhibidor, según la relación siguiente, que depende de la concentración del competidor **I** y de su constante de inhibición, $k_{i,I}$, que, en un sistema competitivo, corresponde a la constante de disociación del competidor **I** ($k_{i,I} = k_{d,I}$) .

$$K'_{D,L} = K_{D,L} \left(1 + \frac{[I]_t}{K_{i,I}} \right) \quad (18)$$

La disminución de la señal de STD del ligando observada, a concentraciones crecientes del inhibidor competitivo **I** ($STD_{L,I}$), se puede describir en función de las constantes de disociación correspondientes al ligando y al inhibidor. La relación entre la saturación observada en presencia de competidor, $STD_{L,I}$, frente a la observada en ausencia de inhibidor, STD_L , permite obtener una expresión independiente de la saturación máxima alcanzable por el ligando ($STD_{L,max}$) :

$$\frac{STD_{L,I}}{STD_L} = \frac{\left(\frac{STD_{L,max} * [L]_t}{K_{D,L} * \left(1 + \frac{[I]_t}{K_{D,I}} \right) + [L]_t} \right)}{\left(\frac{STD_{L,max} * [L]_t}{K_{D,L} + [L]_t} \right)} \quad (19)$$

$$\frac{STD_{L,I}}{STD_L} = \frac{K_{d,L} + [L]_t}{K_{d,L} + \frac{K_{d,L}}{K_{d,I}} [I]_t + [L]_t} \quad (20)$$

El decaimiento del porcentaje de la intensidad de STD de los protones de α -Me-Gal a medida que se van añadiendo concentraciones crecientes de β -Me-Gal se presenta en la figura 4.17.

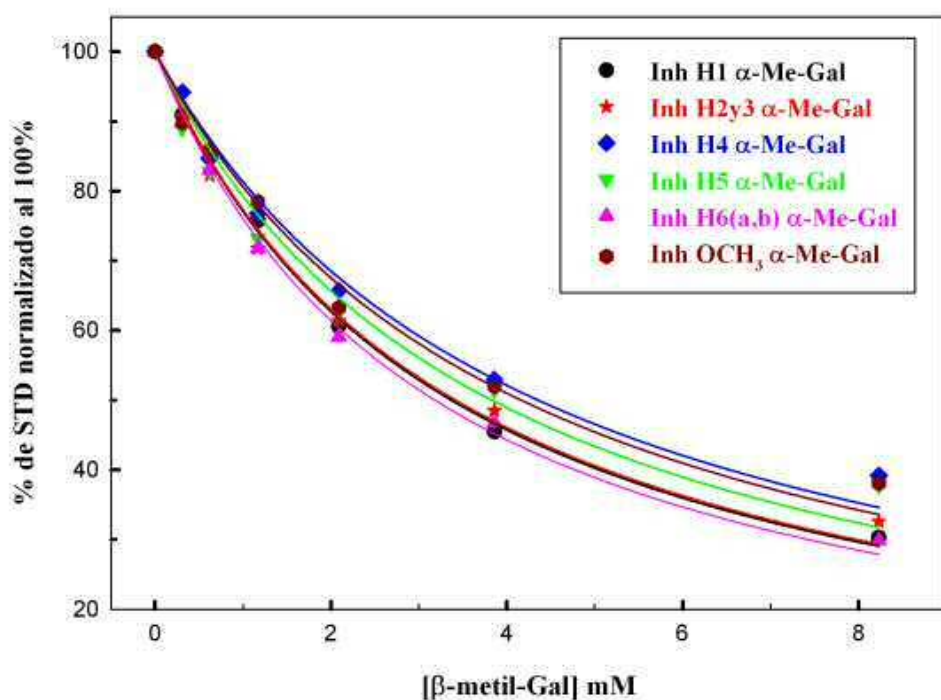


Figura 4.17 Decaimiento de las intensidades de las señales STD de los protones de: la α -Me-Gal a 1 mM con 50 μ M VAA al añadir el inhibidor competitivo β -Me-Gal a concentraciones de 0.3, 0.6, 1.2, 2.1, 3.9, 8.2 mM.

Así se pueden deducir las constantes de afinidad de los protones del α Me Gal y las constantes de inhibición del competidor β Me Gal (Tabla 4.4):

Tabla 4.4 Valores de las constantes de asociación y de inhibición para cada protón determinada a base de la disminución de la intensidad de la señal de los protones de la molécula espía con el aumento de la concentración del inhibidor

Protones	Constante de disociación de la molécula observada (α -Me-Gal)	Constante de inhibición del ligando competidor (β -Me-Gal)
	K_d	K_i
H1	0.56	1.22
H2 y H3	0.56	1.23
H4	0.41	1.27
H5	0.49	1.27
H6(a,b)	0.61	1.21
OCH3	0.44	1.27
Media total	<u>0.52 ± 0.07</u>	<u>1.24 ± 0.03</u>

Analizando las curvas de decaimiento de las señales de saturación de los protones en función del aumento del ligando competidor en el tubo de resonancia, se observó, en ambos casos, que ninguno de los dos ligandos fue capaz de apartar completamente al otro competidor de los sitios de unión. Los valores obtenidos de las constantes de inhibición que ejerce cada ligando sobre el otro (0,52 y 1,24 mM, para el α -Me-Gal y β -Me-Gal, respectivamente) son, aproximadamente, la mitad de los valores obtenidos para las constantes de disociación a partir de las titulaciones directas (1,3 y 2,1 mM, para el α -Me-Gal y β -Me-Gal, respectivamente). Aunque los valores absolutos no coinciden, en cualquier caso, con ambos métodos de medida se observa una preferencia por la unión del α -Me-Gal cercana a un factor de 2. No obstante, en ningún caso se observa la alta selectividad de más de un orden de magnitud descrita para la α -Me-Gal ($K_d = 0,06$ mM) por Bharadwaj et al. [30]. Los valores son más cercanos a los observados como inhibidores de hemaglutinación. Aquí, el α -Me-Gal parece ser 4 veces más eficaz que el β -Me-Gal [39]. Por lo tanto, se vuelve a confirmar que la unión de la VAA para estos monosacáridos no muestra una selectividad significativa. La similitud de la forma de

unión de los ligandos deducida a partir de los perfiles de saturación para cada uno de los protones de ambos monosacáridos podría explicar esa falta de selectividad.

2.3 Elaboración de mapas de epítomos por STD

La intensidad de las señales de STD o la amplitud de saturación expresada en forma porcentual (ecuación 1) se utilizan también para definir el epítomo de unión del ligando[40]. En principio, el porcentaje de saturación de un protón del ligando refleja su proximidad a la superficie de la proteína y se describe generalmente por la ecuación 1 previamente descrita:

$$A_{STD} = \frac{I_0 - I_{sat}}{I_0} \times 100 = \frac{I_{STD}}{I_0} \times 100$$

Durante el periodo de irradiación, la saturación se difunde sobre toda la proteína y se transfiere progresivamente al ligando. Ya se ha mencionado anteriormente los factores a tener en cuenta cuando se analizan cuantitativamente los experimentos STD: tiempos de saturación, la relajación de los espines, el exceso de ligando, etc [32, 41]. Así, por ejemplo, en las tablas 4.5 y 4.6 se observa una gran diferencia en los porcentajes relativos de saturación recibidos por los mismos protones según se realicen los experimentos con tiempos cortos (0,5 s) o tiempos largos (8 s) de saturación.

En el caso de la galactosa, y para ambos anómeros, los protones H4 y H6 sufren mayor saturación usando 0,5 s de irradiación, mientras los protones con tiempos de relajación más largos (H1 en el anómero α y H2 en el β , tabla 4.5) son los que reciben mayor saturación usando tiempos largos, de 8 s. Un comportamiento muy similar se observa en el caso de los α y β metil Gal (tabla 4.6), aunque en el caso del α metil-Gal, la saturación de H2 está promediada con la de H3 por el solapamiento de las señales).

Tabla 4.5 Tiempos de relajación T_1 y valores de STD a tiempos cortos (0,5 s) y largos (8 s) de irradiación (δ 0, 8 ppm) para cada protón de los anómeros α y β Gal (9,8 mM de Gal total, equivalente a 3,3 mM para α -Gal y 6,5 mM para β -Gal) en presencia de viscumina (0,05 mM). En rojo se resaltan los valores máximos de STD y los tiempos de relajación.

<i>α Galactosa</i>					
protón	STD_(0.5s)	%STD_(0.5s)	STD_(8s)	%STD_(8s)	T_1 s
H1	1,74	66,9%	19,41	100%	3.22
H2	1,83	70,4%	16,6	85,5%	2.53
H3	1,85	71,1%	15,97	82,3%	1.95
H4	2,50	96,1%	15,97	82,3%	1.61
H5	1,46	56,1%	12,27	63,21%	1.66
H6	2,60	100%	11,46	59,0%	0.88
<i>β Galactosa</i>					
protón	STD_(0.5s)	%STD_(0.5s)	STD_(8s)	%STD_(8s)	T_1 s
H1	0,70	25%	6,83	33,5%	1.92
H2	1,87	66,8%	20,38	100%	5.34
H3	1,42	50,7%	10,60	52,0%	1.51
H4	2,73	97,5%	14,07	69,0%	1.6
H5	1,48	52,8%	8,07	39,6%	1.02
H6	2,80	100%	11,02	54,1%	0.88

Tabla 4.6 Tiempos de relajación T_1 y valores de STD a tiempos cortos (0,5 s) y largos (8 s) de irradiación (0, 8 ppm) para cada protón de los α y β metil-Gal (7 mM) en presencia de viscumina (0,05 mM). En rojo se resaltan los valores máximos de STD y los tiempos de relajación.

<i>α Metil Galactosido</i>					
protón	STD_(0.5s)	%STD_(0.5s)	STD_(8s)	%STD_(8s)	T_1 s
H1	3,84	53,5%	17,61	77,6%	2,05
H2y3	4,86	67,8%	21,40	94,3%	2,64
H4	6,97	97,2%	22,69	100%	1,9
H5	4,11	57,3%	15,06	66,4%	1,64
H6	7,17	100%	17,10	75,4%	1,02
OCH3	1,76	24,5%	7,83	34,5%	1,52
<i>β Metil Galactosido</i>					
protón	STD_(0.5s)	%STD_(0.5s)	STD_(8s)	%STD_(8s)	T_1 s
H1	2,87	32,3%	14,48	40,7%	1,93
H2	6,26	70,4%	35,55	100%	3,57
H3	5,73	64,4%	22,47	63,2%	1,46
H4	8,81	99,1%	26,78	75,3%	1,48
H5	4,86	54,7%	17,27	48,6%	1,05
H6	8,89	100%	21,61	60,8%	0,88
OCH3	2,25	25,3%	12,03	33,8%	1,48

Este comportamiento diferencial en función del tiempo de saturación está directamente relacionado con los diferentes tiempos de relajación observados para los distintos protones, como dedujeron Krishna y colaboradores [32-33, 42].

Para poder obtener un mapa o epítipo de unión fiable a partir de las intensidades de saturación recibida por los protones del ligando se han desarrollado diferentes estrategias: Shapiro y colaboradores propusieron el empleo de tiempos de saturación relativamente cortos (más cortos que la mitad de T_1) [41]. Por otra parte, Angulo et al. han utilizado la estimación de las velocidades iniciales de crecimiento de los STD a partir del ajuste a una ecuación exponencial creciente (ecuaciones 2 y 3), mediante el registro de curvas de crecimiento a una concentración de ligando y distintos tiempos de saturación. [36-37, 40, 43] Básicamente, es el mismo protocolo que se ha usado en el apartado anterior 4.2.1 para determinar las constantes de afinidad de los ligandos. Recientemente, Claridge et al. [32] han propuesto otra aproximación experimental con el objeto de compensar la influencia de los tiempos de relajación de los protones del ligando sobre la saturación neta observada. Estos autores proponen la introducción de un factor “q”, que representa las tasas relativas de transferencia de saturación a cada núcleo del ligando, corregidas por su correspondiente velocidad de relajación. Experimentalmente, el valor de “q” (ecuación 21) se determina realizando medidas de STD a tiempos largos de saturación, alcanzando el estado estacionario de saturación. Entonces se divide el valor de la saturación observada entre el tiempo de relajación correspondiente a cada núcleo (T_1) del ligando y entre la fracción molar de ligando en forma la unida (χ_{BL}). Esta se estima en condiciones de gran exceso de ligando sobre receptor ($L_0 \gg R_0$), según la siguientes ecuaciones

$$\chi_{BR} = \frac{RL}{R_0} = \frac{L_0}{k_d + L_0} \Rightarrow \chi_{BL} = \frac{RL}{L_0} = \frac{R_0}{k_d + L_0}$$

$$q \approx \frac{f_L \{R\}}{T_1^L * \chi_{BL}} \Rightarrow q \approx \frac{K_d + [L]_0}{[R]_0} * \frac{f_L \{R\}}{T_1^L} \quad (21) \quad \text{Donde}$$

$$f_L\{R\} = \frac{I_0 - I_{SAT}}{I_0} = \frac{I_{STD}}{I_0}$$

$f_L[11]$ es el aumento fraccional o intensidad relativa del STD observado por el ligando cuando se satura el receptor presente en el medio a una concentración definida [11]

A vista de estos antecedentes y antes de pasar a la interpretación estructural de los resultados de STD, fue necesario medir la relajación de cada protón del ligando en ausencia y presencia de viscumina. Mediante el experimento denominado de inversión-recuperación, se estimaron los valores de relajación T_1 correspondientes a cada uno de los protones de la Gal en su mezcla anomérica y de los α y β metil Gal.

Los experimentos de relajación T_1 se realizaron en una espectrómetro de RMN Bruker 500 MHz, a 300 K, usando una relación molar ligando/proteína de aproximadamente 20/1. Los valores de T_1 se obtuvieron del ajuste a una función exponencial creciente del valor de la integral de la señal para cada protón, medido en una serie de espectros en los que se ha varió, entre 0,01 y 10 s, el tiempo de recuperación (τ) entre el pulso inicial de 180° de inversión de la señal y el pulso final de 90° antes de la adquisición:

$$M(\tau) = -M_0 \cdot \exp(-\tau / T_1)$$

Las tablas 4.7 y 4.8 recogen los valores calculados de los tiempos de relajación longitudinal T_1 para cada protón de cada ligando, junto con los valores normalizados de las velocidades iniciales de saturación [36, 41] y del factor q [32]. Estos valores se calcularon a partir de las curvas de crecimiento (V_i) y de las intensidades de STD usando tiempos largos de saturación según la ecuación 21 (q). Los experimentos se realizaron usando concentración alta de ligando (9,8 mM de Gal total, relación carbohidrato: viscumina 66:1 y 130:1 para α -Gal, y β -Gal respectivamente y 7 mM tanto para la α y β -metil-Gal, relación carbohidrato: viscumina 140:1). También, para comparar, se incluyen los valores de $V_{i\max}$ normalizados obtenidos en el apartado 4.2.1, a partir de todas las series de experimentos, combinando las curvas de saturación obtenidas con concentraciones diversas de ligando (figura 4.15, tablas 4.1 y 4.2).

Tabla 4.7 Factores q y velocidades iniciales calculados para la α y β -Gal (relación carbohidrato: viscumina 66:1 y 130:1, respectivamente). Los valores de q fueron estimados utilizando las constantes de disociación calculadas en el apartado 4.2.1. $K_d = 0,72$ mM y $K_d = 1,37$ mM para la α -Gal y la β -Gal, respectivamente). En rojo se resaltan los valores máximos y en amarillo, los mínimos.

<i>α Galactosa</i>						
protón	T_1 s	q /s	q %	V_i /s	V_i %	V_{iMax} %
H1	3.22	4,8	46,3%	0,212	77,8%	68,4%
H2	2.53	5,2	50,3%	0,211	77,5%	64,9%
H3	1.95	6,5	62,9%	0,218	80,1%	82,4%
H4	1.61	7,9	76,2%	0,273	100%	100%
H5	1.66	5,9	56,7%	0,216	79,1%	71,9%
H6	0.88	10,4	100%	0,249	91,4%	91,2%
<i>β Galactosa</i>						
protón	T_1 s	q /s	q %	V_i /s	V_i %	V_{iMax} %
H1	1.92	5,5	28,40%	0,155	29,1%	33,7%
H2	5.34	6,0	30,47%	0,385	72,2%	82,5%
H3	1.51	11,0	56,06%	0,323	60,5%	66,2%
H4	1.6	13,8	70,21%	0,533	100%	100%
H5	1.02	12,4	63,16%	0,274	51,3%	56,2%
H6	0.88	19,7	100%	0,499	93,7%	86,2%

Los epítomos de unión (figuras 4.18 y 4.19) se obtuvieron mediante la normalización porcentual con referencia al valor más alto de la velocidad inicial o del factor q obtenidos. De este modo, las regiones del ligando más cercanas al sitio de reconocimiento del receptor se identifican fácilmente, ya que muestran los valores más altos de saturación.

La figura 4.18 muestra gráficamente los epítomos de unión deducidos para los anómeros de Gal, teniendo en cuenta los rangos relativos porcentuales del factor q obtenido con 8 s de saturación, de la velocidad inicial calculada a partir de las curvas de crecimiento obtenidas con una relación Gal total/viscumina 180:1, y de la velocidad inicial máxima de saturación $V_{i\max}$ para una concentración 50 μ M de viscumina.

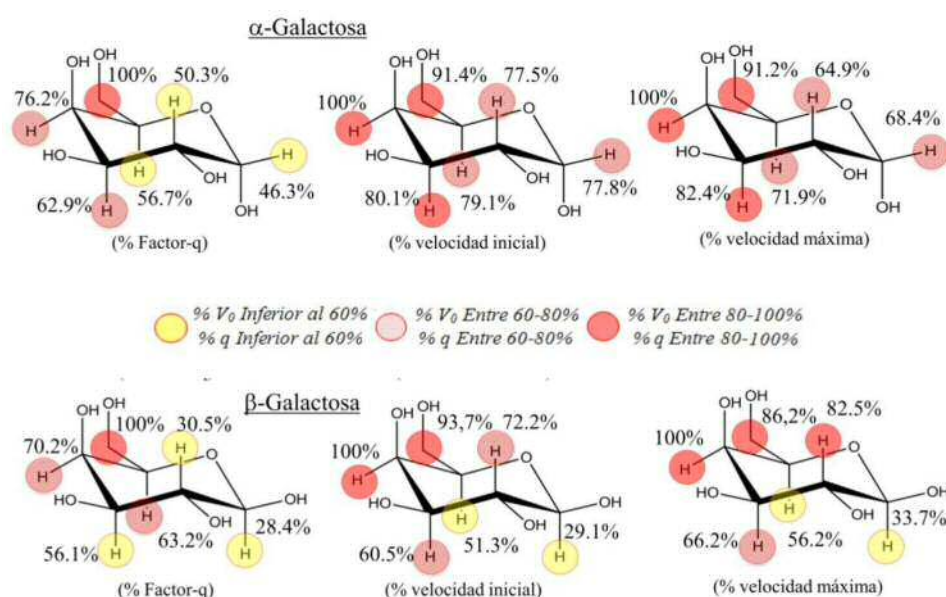


Figura 4.18 Estructura de la α y β -Gal con su grado relativo de saturación de cada protón representado por los porcentajes del factor q, de la velocidad inicial " V_0 " y de la V_{max} , que se normalizan al valor máximo obtenido en cada caso

Se observa que los valores obtenidos, en todos los casos, coinciden cualitativamente en que los H4 de ambos anómeros, α y β -Gal, junto con los H6ab (que en los espectros solapan para ambos anómeros) presentan los valores relativos máximos. Para ambos anómeros, cuando se utiliza el factor q (que prima los protones con tiempos cortos de relajación) para describir el epítipo de contacto con la proteína, son los H6a/b los que alcanzan el 100% del parámetro, seguidos de H4 mientras que, cuando se utilizan las velocidades de saturación, se invierte el orden. Ahora, H4 marca el 100%, seguido de los H6a/b. Por el contrario, H1, prácticamente en todos los casos, presenta los valores relativos más bajos de los dos tipos de parámetros. Por tanto, para ambos anómeros, el epítipo de unión indica una mayor cercanía a la proteína de las posiciones 4 y 6 de la Gal, mientras que la posición anomérica 1 debe quedar más expuesta al disolvente, quedando las posiciones 2, 3 y 5 en una situación intermedia.

En el caso de α y β -Me-Gal, vuelven a ser H6a/b y H4 los que alcanzan los máximos valores relativos, invirtiéndose el orden en algún caso, según sea el factor q, o las velocidades de saturación, el parámetro que se utilice para definir el epítipo. La tabla 4.8 resume los datos obtenidos.

Por otro lado, para α y β -Me-Gal, los porcentajes más bajos tanto del factor q como de la velocidad inicial corresponden a los protones del grupo O-Metilo, seguidos de los protones H1. Estos resultados indican otra vez que la viscumina reconoce mejor la región de H4 y H6 (ab) y que, por el contrario, el sustituyente en la posición anomérica queda más alejado de la superficie de la proteína.

Tabla 4.9 Factores q y velocidades iniciales calculados para los α - y β -Me Gal (relación carbohidrato: viscumina de 170:1). Los valores de q fueron estimados utilizando el promedio de las constantes de disociación e inhibición determinadas en los apartados 4.2.1 y 4.2.2. $K_d = 0,89$ mM para α -Me-Gal y $K_d = 1,69$ mM para β -Me-Gal). En rojo se resaltan los valores máximos y en amarillo, los mínimos.

<i>α Metil Galactosido</i>						
protón	T_1 s	q /s	q %	V_i/s	V_i%	V_{iMax} %
H1	2,05	13.55	51,22%	0,653	47.6%	38,3%
H2y3	2,64	12.79	48,35%	0,881	64.3%	45,9%
H4	1,9	18.84	71.22%	1,261	92,0%	100%
H5	1,64	14.49	54,78%	0,759	55,4%	38,3%
H6	1,02	26.45	100%	1,370	100%	73,9%
OCH3	1,52	8.12	30,69%	0,301	22,0%	13,0%
<i>β Metil Galactosido</i>						
protón	T_1 s	q /s	q %	V_i/s	V_i%	V_{iMax} %
H1	1,93	13.0	30,5%	0,545	33,3%	27,9%
H2	3,57	17.3	40,5%	1,148	70,1%	61,7%
H3	1,46	26.7	62,6%	1,072	65,4%	59,2%
H4	1,48	31.4	73,6%	1,638	100%	100%
H5	1,05	28.6	66,9%	0,930	56,8%	43,2%
H6	0,88	42.7	100%	1,609	98,3%	95,1%
OCH3	1,48	14.1	33,1%	0,375	22,9%	19,7%

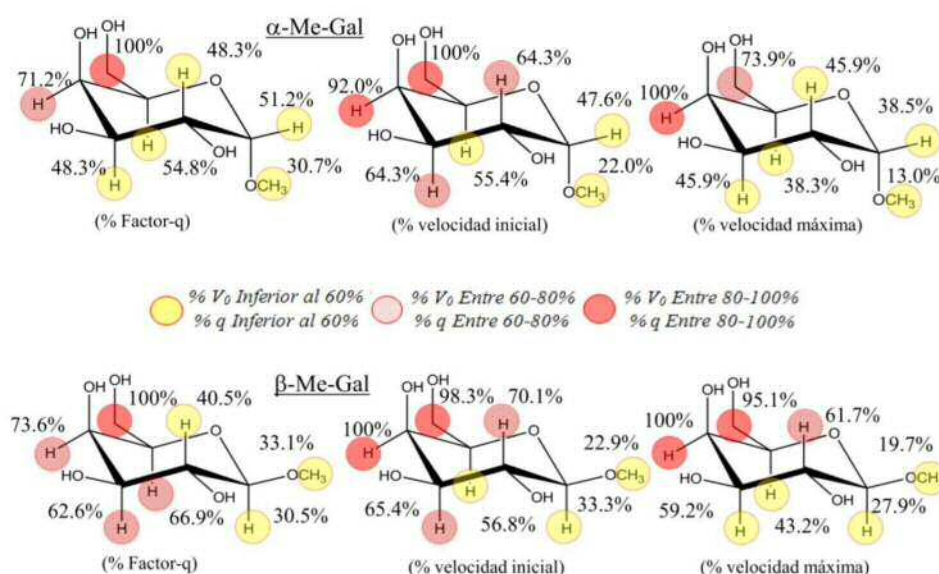


Figura 4.19 Estructura de la α y β -Me Gal con el grado relativo de saturación de cada protón representado por los porcentajes del factor q, de la velocidad inicial " V_0 " y de la V_{max} , que se normalizan al máximo valor obtenido en cada caso.

Estudios previos han mostrado que galactósidos con ambas configuraciones, α y β , son reconocidos por la VAA [31, 39, 44]. Por otro lado, en las estructuras cristalográficas de los complejos de viscumina con galactosa (pdb: 1PUM) y lactosa (pdb: 1PUU) [27] se observa que todos los grupos OH del residuo de Gal participan en diferentes grados en la interacción con viscumina. El grupo 4'OH, que define la estereoquímica de la galactosa, es el que parece más importante, seguido por el 3'OH, en concordancia con la especificidad epimérica por ligando que presenta la viscumina.

A continuación, para poder trasladar la descripción de epítomos de contacto a un modelo estructural del modo de unión de la Gal y los metil Gal a la viscumina se procedió a cálculos de acoplamiento o encaje intermolecular de ligando "*docking*" con ambos anómeros de Gal y de sus metil derivados. A partir de las coordenadas de los modelos estructurales obtenidos se pueden predecir, teóricamente, las transferencias de saturación a cada protón del ligando y compararlas con los valores de STD experimentales utilizando el protocolo CORCEMA, desarrollado por Krishna y colaboradores [32-33, 42].

2.4 Cálculos de "Docking" y CORCEMA

En la estructura del complejo con galactosa (código PDB: 1PUM) [27], la Gal está en forma β en los dos sitios de unión ($\alpha 1$, definido por W38 y $\gamma 2$, definido por Y249, véase el apartado 4.1.1, figuras 4.2 y 4.3 respectivamente). Esta estructura se utilizó como estructura de partida para generar modelos moleculares mediante cálculos de "docking" de los cuatro ligandos (α - y β -Gal y α - y β -metil Gal) interaccionando con los dos sitios de unión a carbohidratos de la viscumina. La estructura 1PUM se optimizó para estos cálculos usando las herramientas de preparación de biomoléculas incluidas en el programa Maestro del paquete Schrödinger (Maestro, versión 9.0, Schrödinger, LLC, New York, NY, 2009). Los cálculos de "docking" se realizaron utilizando el programa AutoDock [45-46]. Se preparó una caja reticulada cúbica ("*grid-box*") con 70 puntos para cada dimensión y un espaciado de 0.325 Å entre cada punto, que se centró usando como referencia la posición del O5 de las β -Gal cristalizadas en cada uno de los dos sitios de reconocimiento de Gal (figuras 4.2 y 4.3). El cálculo se realizó asumiendo un receptor rígido e incluyendo flexibilidad en el ligando. Se permitieron rotaciones en los ángulos de torsión del grupo hidroximetilo y de los grupos hidroxilos. En el caso de los metil Gal, también se permitió la rotación del enlace glicosídico sumando hasta un total de 6 grados de libertad de torsión. Se realizaron 100 cálculos usando un algoritmo genético Lamarckiano y las opciones restantes se dejaron por defecto. Después del cálculo, se realizó un agrupamiento ("clustering") con un criterio de RMSD (desviación de la raíz cuadrática media de las coordenadas) de 2 Å.

Las soluciones fueron elegidas teniendo como criterio la energía de unión o "Docking", así como el acuerdo con los datos experimentales de difracción de Rayos X. Para los otros ligandos, se generaron manualmente modelos geométricos a partir de las estructuras con β -Gal, simplemente cambiando la estereoquímica de la posición anomérica para la α -Gal y añadiendo el grupo metilo, en ambas configuraciones, para los modelos de α - y β - metil-Gal.

En el "docking" de la α -Gal al sitio $\alpha 1$ (W38, figura 4.20, izquierda), el programa reprodujo, en el 3º clúster en términos de energía de unión, una geometría cercana a la del modelo derivado de la estructura cristalográfica, presentando los protones H3, H4, H5 y H6a/b hacia la cadena lateral de W38. El primer clúster, con energía más favorable, mostró una orientación del anillo de piranosa rotado aproximadamente 180°

alrededor de un eje que atraviesa los enlaces O5-C5 y C2-C3. En el segundo clúster (el más poblado), la orientación de la Gal mostraba una rotación alrededor de un eje ortogonal al anterior y perpendicular al plano de la piranosa definido por los átomos C2-C3-O5-C5. En el caso del sitio $\gamma 2$ (Y249, figura 4.20 derecha), también se seleccionaron los 3 primeros clústeres en términos energéticos. El 1º y el 3º presentaban los protones H3, H4, H5 y H6a/b hacia la cadena lateral de Y249, aunque el anillo de piranosa estaba, en el primer caso, orientado casi ortogonal al anillo de la tirosina. En el 3º clúster, ambos anillos están casi paralelos. Esta orientación está favorecida en el "docking" por el establecimiento de un enlace de hidrógeno entre el OH anomérico axial de la α -Gal y el hidroxilo fenólico de la tirosina). En el modelo derivado de la estructura cristalográfica, el anillo de piranosa presenta una inclinación intermedia entre la de los clústeres 1º y 3º, respecto a la tirosina. El segundo clúster obtenido para el sitio $\gamma 2$ es el más poblado. Aquí, el anillo de piranosa estaba rotado aproximadamente 180º alrededor de un eje sobre el plano de la piranosa, atravesando los enlaces O5-C5 y C2-C3. La Gal presenta los protones H1 y H2 al anillo fenólico de la tirosina.

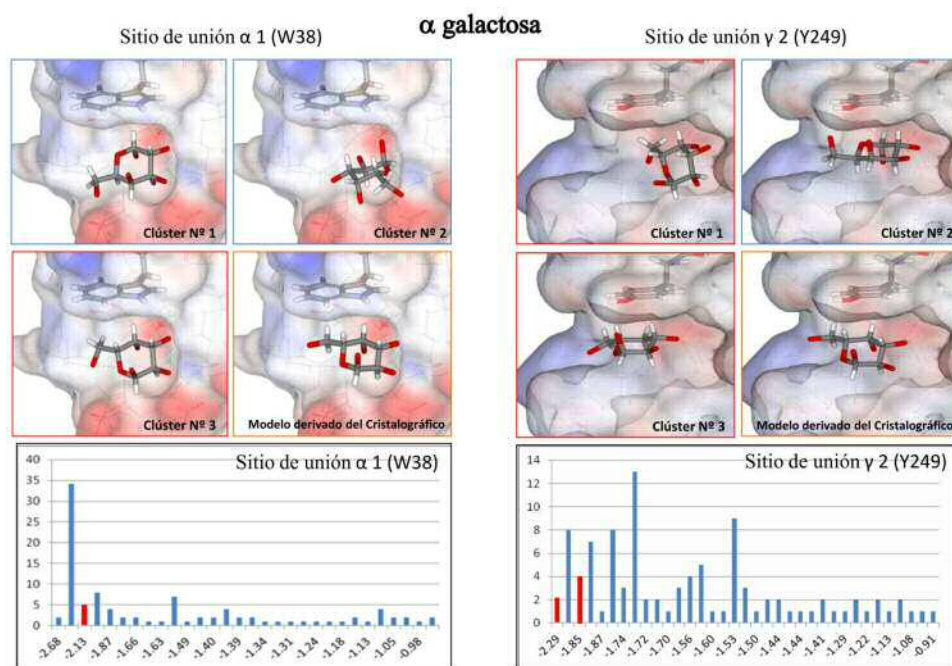


Figura 4.20 Representación de la distribución de clústeres y geometrías seleccionadas del cálculo con AUTODOCK del modo de unión de la α -Gal en el sitio $\alpha 1$ (izquierda) y en el sitio $\gamma 2$ (derecha). Se incluye el correspondiente modelo derivado de la geometría de la β -Gal presente en la estructura IPUM para ambos sitios (recuadrado en naranja). En rojo se resaltan los correspondientes clústeres con la geometría más cercana al modelo.

En el caso de la β -Gal, los dos primeros clústeres sobre el sitio $\alpha 1$ (figura 4.21 izquierda) reproducían la orientación de la estructura cristalográfica. En cambio, en el "docking" sobre el sitio $\gamma 2$ (figura 4.21 derecha) el primer clúster (con energía de unión más favorable, aunque muy poco poblado) es el que se más acercaba a la geometría de la estructura cristalográfica. Los clústeres 2º y 3º sobre el sitio $\gamma 2$, aunque más poblados, presentaban al anillo de piranosa paralelo al anillo de tirosina pero rotado 180° y enfrentando a éste únicamente el H2, situado en la cara β de la piranosa.

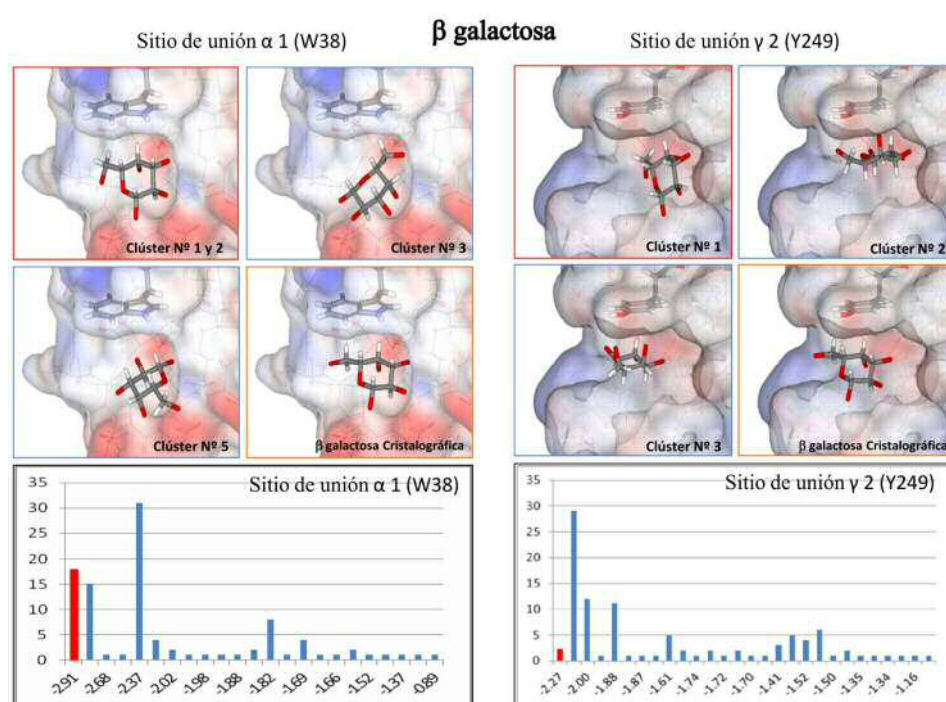


Figura 4.21 Representación de la distribución de clústeres y geometrías seleccionadas del cálculo con AUTODOCK del modo de unión de la β -Gal en el sitio $\alpha 1$ (izquierda) y en el sitio $\gamma 2$ (derecha). Se incluye el correspondiente modelo derivado de la geometría de la β -Gal presente en la estructura IPUM para ambos sitios (recuadrado en naranja). En rojo se resaltan los correspondientes clústeres con la geometría más cercana al modelo.

En el caso del α -metil-Gal (figura 4.22), el "docking" sobre la posición $\alpha 1$ rindió prácticamente un único clúster, muy poblado, con la energía de unión más favorable que reproducía la geometría del modelo de unión de la estructura cristalográfica. En el "docking" sobre la posición $\gamma 2$ se seleccionaron 3 clústeres, el 1º muy poco poblado y el 3º el más poblado, ambos con una geometría cercana a la derivada de la cristalográfica.

La Gal presenta los protones H3, H4, H5, H6a/b orientados hacia la tirosina, mientras que el 2º clúster tenía el anillo de piranosa rotado 180° presentando la cara β de la piranosa, con los protones H1 y H2, hacia el anillo de tirosina.

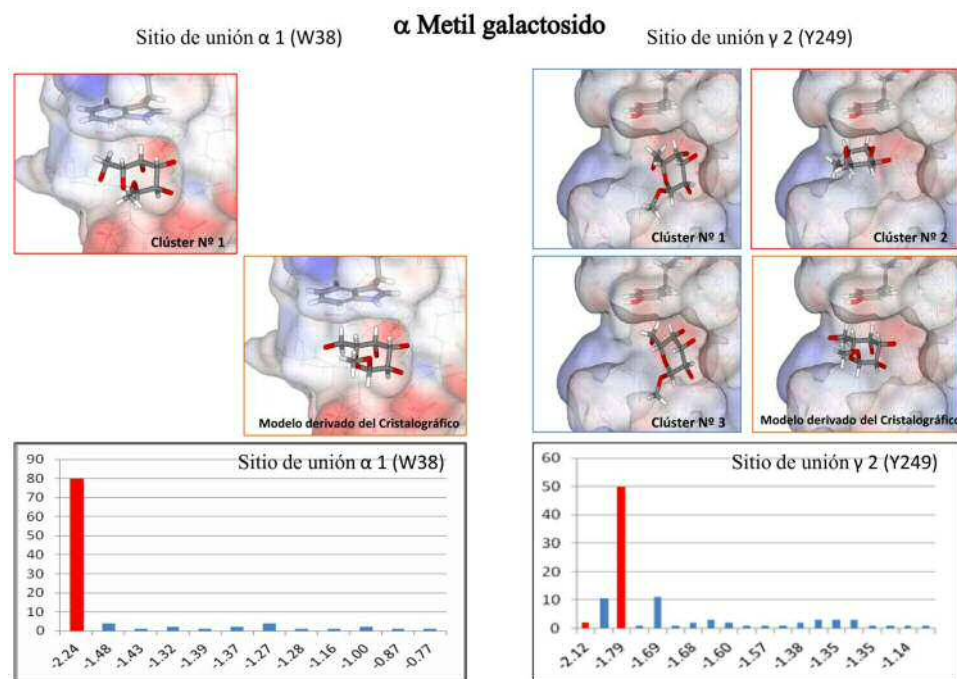


Figura 4.22 Representación de la distribución de clústeres y geometrías seleccionadas del cálculo con AUTODOCK del modo de unión de la α -Me Gal en el sitio α 1 (izquierda) y en el sitio γ 2 (derecha). Se incluye el correspondiente modelo derivado de la geometría de la β -Gal presente en la estructura IPUM para ambos sitios (recuadrado en naranja). En rojo se resaltan los correspondientes clústeres con la geometría más cercana al modelo.

En el caso del β -metil-Gal (figura 4.23), el "docking" sobre la posición α 1 rindió un clúster muy poblado, con la energía de unión más favorable, que reproducía la geometría del modelo de unión derivado de la estructura cristalográfica. En el "docking" sobre la posición γ 2 se seleccionaron 2 clústeres. El primero mostró una geometría cercana a la derivada de la estructura cristalográfica, presentando hacia la tirosina los protones H3, H4, H5, H6a/b. En el segundo clúster, el anillo de piranosa vuelve a estar rotado 180° presentando la cara β de la piranosa, el H2, hacia el anillo de tirosina.

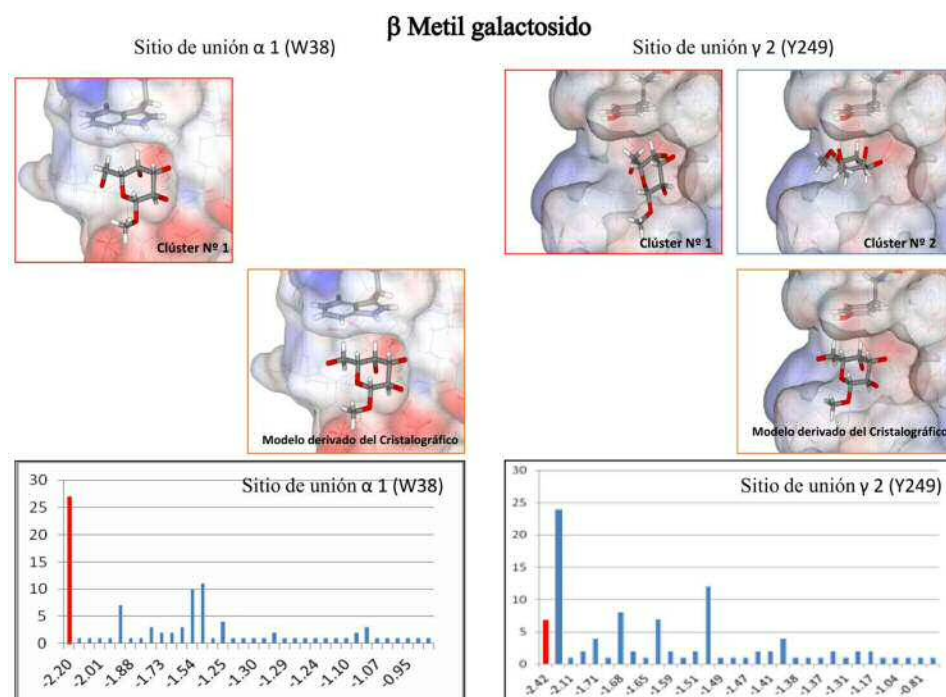


Figura 4.23 Representación de la distribución de clústeres y geometrías seleccionadas del cálculo con AUTODOCK del modo de unión de la β -Me Gal en el sitio α 1 (izquierda) y en el sitio γ 2 (derecha). Se incluye el correspondiente modelo derivado de la geometría de la β -Gal presente en la estructura 1PUM para ambos sitios (recuadrado en naranja). En rojo se resaltan los correspondientes clústeres con la geometría más cercana al modelo.

Para los cuatro galactósidos, los cálculos de "docking" realizados con Autodock permitieron encontrar modelos estructurales con posturas del correspondiente ligando en geometrías cercanas a la de la estructura cristalográfica de los complejos de la β -galactosa en los dos sitios de unión de la viscumina. Un motivo común de todas estas posturas es, por un lado, la presentación de los protones H3, H5 y especialmente el protón H4 y uno de los protones H6 orientados hacia el anillo indólico del triptófano W38 (sitio α 1) y hacia el anillo fenólico de la tirosina Y249 (sitio γ 2). Por otro lado, la posición anomérica se orienta hacia la parte externa del sitio de unión. Cualitativamente, este modo de unión es compatible con los epítopos deducidos mediante STD. En estos, para los 4 compuestos, se encuentra que los protones H4 y H6 reciben la máxima saturación y deben, por tanto, estar muy cercanos a la proteína. En cambio, H1 y el grupo metilo de los metil Gal son los que reciben menor saturación y, previsiblemente, deben quedar situados en la zona del sitio de unión más expuesta al disolvente.

Para validar los modelos tridimensionales de los complejos obtenidos a partir de los cálculos de "docking" se procedió a realizar una predicción teórica, mediante el protocolo CORCEMA, de las transferencias de saturación desde la proteína a los protones de los ligandos para cada una de las geometrías seleccionadas en los cálculos de docking (figuras 4.20-23). A continuación, los valores obtenidos se compararon con los resultados experimentales de STD.

Los cálculos de CORCEMA se realizaron con cada una de las posturas de los clústeres seleccionados. En todos los casos, se definió un radio de 10 Å alrededor del ligando. Todos los protones no intercambiables de la proteína y del ligando dentro de esta región se usaron para calcular la matriz de relajación completa. Seguidamente se introdujeron en el programa los valores de tiempo de irradiación, de las concentraciones de ligando y receptor, de la constante de afinidad correspondiente (calculadas en los apartados 4.2.1 y 4.2.2) y del tiempo de correlación de la viscumina. Se estimó un valor de 70 ns para este parámetro, usando el programa HYDROPRO [47], considerando que la viscumina, a la concentración usada (0,05 mM), está presente en disolución como un heterotetrámero ABAB [12, 48]. La irradiación en la zona alifática del espectro se simuló considerando directamente saturados todos los protones de los grupos metilo de las cadenas laterales de los residuos de la proteína.

En la figura 4.24 se muestran, para cada ligando y cada sitio de unión ($\alpha 1$ y $\gamma 2$), los perfiles de los porcentajes de saturación (2 s) de los protones del ligando. También se muestra el perfil de saturación obtenido experimentalmente (línea azul) en los espectros STD realizados con 2 segundos de irradiación a δ 0,8 ppm.

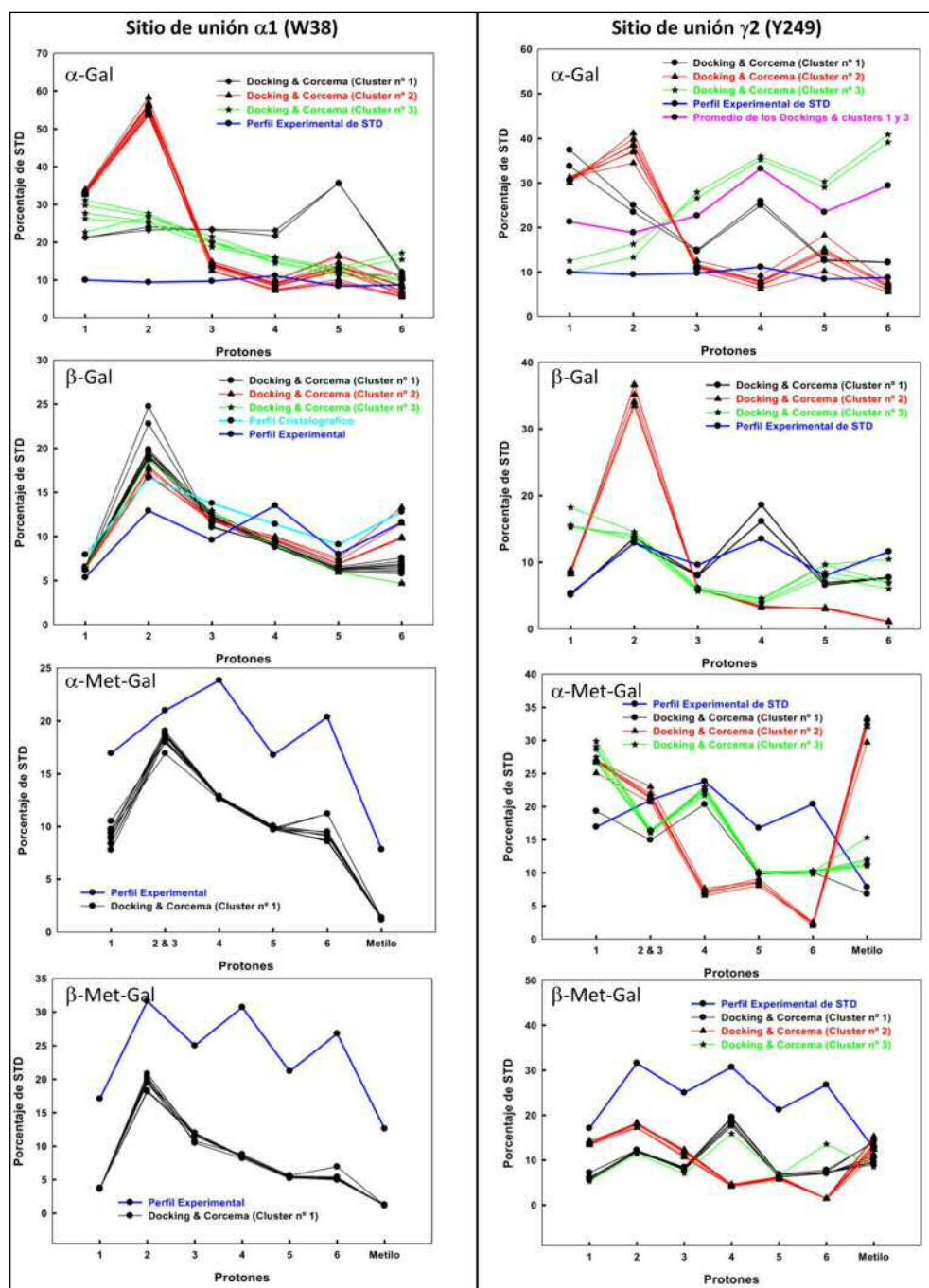


Figura 4.24 Perfiles de STD de los protones de los cuatro ligandos, calculados usando CORCEMA sobre los modelos estructurales obtenidos por cálculos de "docking" con AUTODOCK en los dos sitios de unión de la viscumina. Los resultados para el sitio $\alpha 1$ se describen en los paneles de la izquierda. Los del sitio $\beta 2$ en los paneles de la derecha. Los clústeres se corresponden con los descritos en las figuras 4.20-4.23. En azul se representan los perfiles de saturación experimentales para cada ligando obtenidos con 2 segundos de irradiación (δ 0,8 ppm) en presencia de 0,0 5mM de viscumina

Es interesante resaltar que, para todos de los ligandos, los clústeres seleccionados para el sitio de unión $\alpha 1$ (W38, figura 24, paneles de la izquierda) nunca reproducen el perfil de saturación observado experimentalmente (línea azul). En el caso de la α -Gal, H4 recibe experimentalmente la mayor saturación, mientras que las geometrías seleccionadas dan mayor saturación a H2 en los clústeres 2 y 3, o a H5 en el clúster 1. En el caso de la β -galactosa ocurre algo similar. Los clústeres 1 y 2, cercanos a la geometría de la estructura cristalina, así como la propia estructura cristalina (línea turquesa), predicen una mayor saturación de H2. En cambio, el perfil experimental de saturación tiene el máximo sobre H4, seguido por H2 y H6. En el caso de los metil-Gal, en ambos casos, el "docking" rindió un clúster muy mayoritario, con una geometría cercana a la del modelo cristalográfico. El cálculo de CORCEMA indicó un máximo de saturación sobre H2, mientras que los perfiles experimentales mostraron picos de saturación en los protones H4 y H6 y también sobre H2 en el caso del β -metil-galactósido. En el caso del α -metil Gal, el solapamiento de las señales de H2 y H3 no permite discriminar la saturación para cada protón.

Por el contrario, cuando se analizan los resultados de las predicciones de saturación de CORCEMA sobre los clústeres seleccionados para el sitio $\gamma 2$ de los ligandos con configuración β (β -Gal y β -metil Gal, figura 4.24, 2° y 4° paneles de la columna derecha) se observa una clara concordancia entre el perfil de saturación experimental (líneas azules) y la saturación calculada para los clústeres cercanos a la geometría del modelo cristalográfico (clúster 1 en el caso de la β -Gal y clústeres 1 y 3 en el caso del β -metil Gal). Los picos de saturación se dan en los protones H2, H4 y H6. En el caso de los ligandos con configuración α , la correlación entre los perfiles de saturación experimentales y los calculados es incompleta. En el caso del α -metil Gal (figura 4.24, 3° panel de la derecha), los clústeres 1 y 3 muestran la geometría más cercana al modelo cristalográfico, aunque presentan una orientación ortogonal entre el anillo de piranosa y el anillo aromático de la tirosina. Los cálculos coinciden con el resultado experimental (una alta saturación sobre H4), pero sobreestiman la saturación observada sobre H1. La orientación ortogonal de los anillos y la configuración α sitúan a H1 directamente orientado hacia el interior del sitio de unión de la proteína (figura 4.22 panel derecho) e infravaloran la saturación recibida por los protones H6. En el caso de la α -Gal (figura 4.24, primer panel de la derecha), tampoco ningún clúster de manera individual

reproduce el perfil de saturación experimental. Sin embargo, los clústeres 1 y 3 (figura 4.20, panel derecho) muestran geometrías cercanas a la del modelo cristalográfico, pero con orientaciones relativas de los anillos de piranosa y fenol en disposiciones extremas, ortogonal y paralela respectivamente. Los cálculos proporcionan, en ambos casos, una alta saturación sobre H4. Su combinación lineal (figura 4.24, primer panel de la derecha, línea fucsia) reproduce el perfil de saturación experimental (línea azul), aunque no los valores absolutos.

Finalmente, hay que mencionar que, para los cuatro ligandos, todos aquellos clústeres seleccionados que se alejan de la geometría derivada del modelo cristalográfico nunca reproducen los perfiles de saturación experimentales.

3 DISCUSSION

La viscumina es una lectina específica de galactósidos que reconoce oligosacáridos que contienen galactosa en su extremo no reductor [27, 30-31, 39]. Aunque ambas configuraciones anoméricas, α y β , son reconocidas, en casos particulares, como en el caso de los metil-galactósidos, se ha descrito una significativa mayor afinidad hacia el isómero con configuración α [30, 39]. Sin embargo, en el caso más sencillo de una pareja de anómeros, como es la propia galactosa, únicamente existen datos globales de afinidad de la viscumina por la mezcla de ambos anómeros ($K_d=1$ mM, [30]). Obviamente, dadas las características propias de una mezcla anomérica en equilibrio, no se pudo deducir si existía selectividad hacia alguna de las dos formas. En esta Tesis, mediante experimentos de STD-RMN ha sido posible estudiar, de manera individualizada, la interacción de cada uno de los anómeros de la Gal con la viscumina, aun cuando están presentes en una mezcla en equilibrio termodinámico. El análisis de las curvas de crecimiento de STD junto con la derivación de las ecuaciones de equilibrio apropiadas para interpretar el comportamiento de una mezcla de ligandos alternativos mutuamente competitivos y presentes en una relación constante (ecuaciones 12 y 13), ha permitido extraer los valores de constantes de disociación individuales para cada anómero de la Gal. Los resultados obtenidos (tabla 4.3) muestran que la viscumina tiene una mayor preferencia por el anómero α -Gal ($K_d=0,7\pm0,3$ mM) que por el β -Gal ($K_d=1,4\pm0,5$ mM), aunque la precisión de las medidas minimiza la significación de la diferencia observada. Por otro lado, usando galactosas bloqueadas en la posición

anomérica (α - y β -metil Gal), se había descrito una selectividad de un orden de magnitud a favor del isómero α ($K_d = 0,06$ mM para el α -metil-Gal y $K_d = 0,9$ mM para el β -metil-Gal [30]). Sin embargo, al estudiar ambos isómeros metilados de la galactosa con la misma metodología de STD-RMN usada para la galactosa, los resultados obtenidos solamente reflejan una diferencia de selectividad de un factor de 2 a favor del α -Me-Gal ($K_d = 1,3 \pm 0,7$ mM y $2,1 \pm 0,8$ mM para el α - y β -Me-Gal, respectivamente). Esta baja selectividad se confirmó realizando experimentos de STD-RMN de competición mutua entre ambos isómeros. De nuevo se obtuvieron constantes de disociación diferenciadas con un factor de 2 en favor del isómero α (tabla 4.4). En este punto, hay que tener en cuenta que la viscumina tiene dos sitios de unión a carbohidratos (sitio $\alpha 1$, asociado al residuo W38 y sitio $\gamma 2$, asociado al residuo Y249) [27]. Solís et al. [48] han propuesto que estos sitios son de alta y baja afinidad, respectivamente, mostrando que la funcionalidad de cada uno de los sitios de unión depende del estado oligomérico de la lectina. De hecho, cuando la viscumina, que es un heterodímero con dos cadenas A y B, está presente en disolución en forma de dímero de heterodímeros (ABAB), se ha comprobado que el sitio $\alpha 1$ (W38) tiene una funcionalidad reducida. De hecho, este sitio queda tan cerca de la interfaz de dimerización que se dificulta la entrada de los ligandos [48]. El estado de oligomerización es dependiente de la concentración. Por encima de $38 \mu\text{M}$, se ha mostrado la presencia preferente de la forma ABAB [12]. Los experimentos de STD-RMN del presente trabajo se han realizado usando viscumina con una concentración de $50 \mu\text{M}$ y, por tanto, debe estar mayoritariamente en forma de dímero de heterodímeros ABAB. En este contexto, los resultados de las curvas de crecimiento de STD se trataron de ajustar considerando la presencia tanto de dos sitios de unión independientes, como de un único sitio de unión. Aunque se obtuvieron mejores ajustes de los datos experimentales con el modelo de un único sitio de unión, fue necesario introducir un término adicional inespecífico en las ecuaciones correspondientes a las isotermas de unión (ecuaciones 14, 15 y 17). Este término puede englobar otros procesos, como la presencia del equilibrio de disociación del heterotetrámero y/o una ocupación parcial del otro sitio de unión. Adicionalmente, los experimentos de inhibición mutua, realizados con la misma concentración de viscumina, también mostraron una diferencia de afinidad de sólo un factor de 2 en favor del isómero α -Me Gal. P

or otro lado, los experimentos de STD-RMN, además de posibilitar el cálculo de afinidades [36], también contienen información estructural y permiten definir los epítomos de contacto con la proteína desde el punto de vista del ligando [40, 49]. La información experimental de los epítomos calculados a partir de los experimentos STD-RMN se puede transformar en una interpretación estructural mediante la construcción de modelos estructurales, usando cálculos de "docking". Con estos modelos estructurales fue posible predecir las saturaciones teóricas de los protones del ligando mediante el protocolo CORCEMA, que considera una matriz de relajación completa con los espines del ligando y de la proteína [33-35]. La comparación de las predicciones teóricas con los datos experimentales de saturación permitió seleccionar, entre todas las posturas del ligando en el sitio de unión obtenidas en los cálculos de "docking", aquellas que mejor reprodujeron los valores experimentales. Realizando este análisis para los 4 ligandos en cada uno de los dos sitios de unión (figuras 4.20-23) se dedujo que, en todos los casos, las posturas obtenidas en el sitio de unión $\alpha 1$ (W38) no pudieron explicar los resultados experimentales. Por el contrario, algunas posturas generadas en el sitio $\gamma 2$ (Y249) para los cuatro ligandos (o combinaciones de ellas en el caso de la α -Gal) correlacionaron adecuadamente con los perfiles experimentales de saturación observados, aunque los valores absolutos de saturación no quedaron bien reproducidos. En todos los casos, las posturas que mejor reprodujeron los resultados experimentales fueron aquellas que situaron el anillo de piranosa de la galactosa en una orientación cercana a la observada en la estructura cristalográfica 1PUM del complejo galactosa-viscumina. En esta orientación, los protones H3, H4, H5 y H6 apuntan hacia el anillo aromático de la tirosina Y249.

Estos resultados, junto con los cálculos de constantes de disociación, son consistentes con la interpretación de que, en las condiciones experimentales usadas en este estudio, los dos anómeros de Gal, así como sus metil derivados se unen preferentemente al sitio $\gamma 2$ (Y249) de la viscumina con una constante de asociación en el rango milimolar y con una selectividad favorable, con un factor de 2, hacia los isómeros en configuración α . Para poder estudiar el modo de unión en el sitio $\alpha 1$ (W38) sería necesario cambiar las condiciones experimentales. La reducción de la concentración de proteína debería favorecer la presencia de las formas heterodiméricas AB, evitando el bloqueo del sitio $\alpha 1$, que se produce en la forma ABAB.

4 CONCLUSIONES PARA EL CAPITULO 4

Se ha puesto a punto una metodología basada en experimentos de STD-NMR que permite caracterizar, de manera individual, la afinidad de cada uno de los anómeros del monosacárido galactosa, frente a la lectina de origen vegetal, viscumina, extraída del muérdago. El protocolo ha permitido usar la mezcla de ambos anómeros, presentes en equilibrio termodinámico en disolución.

Se ha mostrado que la viscumina, cuando está presente en concentraciones que favorecen la presencia de dímeros de heterodímeros, es capaz de unir ambos anómeros α y β de galactosa, así como sus correspondientes derivados α -y β -metil galactósidos. Las constantes de disociación están en el rango milimolar y muestra una preferencia cercana a un factor de 2 hacia los isómeros con configuración α .

Mediante la generación de modelos estructurales usando cálculos de "docking" ha sido posible predecir teóricamente la saturación de cada uno de los protones de los cuatro ligandos cuando están unidos a la viscumina, empleando el programa CORCEMA. La comparación de los resultados experimentales de STD con los valores calculados teóricamente muestran que, bajo las condiciones experimentales usadas, los dos anómeros de Gal, así como sus metil derivados, se unen preferentemente al sitio $\gamma 2$ (Y249) de la viscumina. Los modos de unión sitúan el anillo de piranosa en una orientación cercana a la observada en la estructura cristalográfica 1PUM del complejo Gal-viscumina.

5-REFERENCIAS BIBLIOGRAFICAS

1. G.G. Bouffard, K.E. Rudd y S.L. Adhya, *Dependence of lactose metabolism upon mutarotase encoded in the gal operon in Escherichia coli*. Journal of Molecular Biology, 1994. **244**(3): p. 269-278.
2. S.J. Benkovic, *Anomeric specificity of carbohydrate-utilizing enzymes*. Methods in enzymology, 1979. **63**: p. 370-379.
3. J. Salas, M. Salas, E. Vinuela y A. Sols, *Glucokinase of Rabbit Liver*. The Journal of biological chemistry, 1965. **240**: p. 1014-1018.
4. M.J. Borrok, L.L. Kiessling y K.T. Forest, *Conformational changes of glucose/galactose-binding protein illuminated by open, unliganded, and ultra-high-resolution ligand-bound structures*. Protein Sci, 2007. **16**(6): p. 1032-1041.
5. K. Gehring, P.G. Williams, J.G. Pelton, H. Morimoto y D.E. Wemmer, *Tritium NMR spectroscopy of ligand binding to maltose-binding protein*. Biochemistry, 1991. **30**(22): p. 5524-5531.
6. S. Andre, F. Javier Canada, T. Shiao, L. Largartera, T. Diercks, S. André, F.J. Cañada, M. Bergeron Brlek, K. el Biari, A. Papadopoulos, J. Ribeiro, M. Touaibia, D. Solís, M. Menéndez, J. Jiménez Barbero, R. Roy y H.-J. Gabius, *Fluorinated Carbohydrates as Lectin Ligands: Biorelevant Sensors with Capacity to Monitor Anomer Affinity in 19F-NMR-Based Inhibitor Screening*. European journal of organic chemistry, 2012. **2012**(23): p. 4354-4364.
7. H. Franz, P. Ziska y A. Kindt, *Isolation and properties of three lectins from mistletoe (Viscum album L.)*. Biochem J, 1981. **195**(2): p. 481-484.
8. J.B. Dietrich, G. Ribereau-Gayon, M.L. Jung, H. Franz, J.P. Beck y R. Anton, *Identity of the N-terminal sequences of the three A chains of mistletoe (Viscum album L.) lectins: homology with ricin-like plant toxins and single-chain ribosome-inhibiting proteins*. Anticancer Drugs, 1992. **3**(5): p. 507-511.
9. A. Pusztai, S.W. Ewen, G. Grant, W.J. Peumans, E.J. Van Damme, M.E. Coates y S. Bardocz, *Lectins and also bacteria modify the glycosylation of gut surface receptors in the rat*. Glycoconj J, 1995. **12**(1): p. 22-35.
10. L. Barbieri, M.G. Battelli y F. Stirpe, *Ribosome-inactivating proteins from plants*. Biochim Biophys Acta, 1993. **1154**(3-4): p. 237-282.
11. Agapov, I.I., A.G. Tonevitsky, M.M. Moysenovich, N.V. Maluchenko, R. Weyhenmeyer y M.P. Kirpichnikov, *Mistletoe lectin dissociates into catalytic and binding subunits before translocation across the membrane to the cytoplasm*. FEBS Lett, 1999. **452**(3): p. 211-214.
12. M. Jimenez, J.L. Saiz, S. Andre, H.J. Gabius y D. Solis, *Monomer/dimer equilibrium of the AB-type lectin from mistletoe enables combination of toxin/agglutinin activities in one protein: analysis of native and citraconylated proteins by ultracentrifugation/gel filtration and cell biological consequences of dimer destabilization*. Glycobiology, 2005. **15**(12): p. 1386-1395.
13. S. Andre, P.J. Ortega, M.A. Perez, R. Roy y H.J. Gabius, *Lactose-containing starburst dendrimers: influence of dendrimer generation and binding-site orientation of receptors (plant/animal lectins and immunoglobulins) on binding properties*. Glycobiology, 1999. **9**(11): p. 1253-1261.
14. S. Andre, R.J. Pieters, I. Vrasidas, H. Kaltner, I. Kuwabara, F.T. Liu, R.M. Liskamp y H.J. Gabius, *Wedglike glycodendrimers as inhibitors of binding of mammalian galectins to glycoproteins, lactose maxicusters, and cell surface glycoconjugates*. Chembiochem, 2001. **2**(11): p. 822-830.
15. D. Gupta, H. Kaltner, X. Dong, H.J. Gabius y C.F. Brewer, *Comparative cross-linking activities of lactose-specific plant and animal lectins and a natural lactose-binding*

- immunoglobulin G fraction from human serum with asialofetuin*. Glycobiology, 1996. **6**(8): p. 843-849.
16. A.V. Timoshenko, I.V. Gorudko, H. Kaltner y H.J. Gabius, *Dissection of the impact of various intracellular signaling pathways on stable cell aggregate formation of rat thymocytes after initial lectin-dependent cell association of using a plant lectin as model and target-selective inhibitors*. Mol Cell Biochem, 1999. **197**(1-2): p. 137-145.
17. W. Dettmann, M. Grandbois, S. Andre, M. Benoit, A.K. Wehle, H. Kaltner, H.J. Gabius y H.E. Gaub, *Differences in zero-force and force-driven kinetics of ligand dissociation from beta-galactoside-specific proteins (plant and animal lectins, immunoglobulin G) monitored by plasmon resonance and dynamic single molecule force microscopy*. Arch Biochem Biophys, 2000. **383**(2): p. 157-170.
18. V.S. Pashkov, Agapov, I.I., T.A. Balashova, N.S. Egorova, A. Surovoi, I.B. Pevzner y A.G. Tonevitskii, *[Epitope of the catalytic subunit of viscumin, exposed on its interaction with a lipid bilayer]*. Bioorg Khim, 2003. **29**(6): p. 589-596.
19. J.M.R. Lord, L. M., *The intracellular transport of ricin : why mammalian cells are killed and how Ricinus cells survive : Targeting and glycosylation of plant secretory proteins*, in *SFPV : French Society of Plant Physiology meeting*, Elsevier, Editor. 1996, 1995-03-31: Bois-Guillaume, FRA. p. 253-261.
20. B. Mockel, T. Schwarz, H. Zinke, J. Eck, M. Langer y H. Lentzen, *Effects of mistletoe lectin I on human blood cell lines and peripheral blood cells. Cytotoxicity, apoptosis and induction of cytokines*. Arzneimittelforschung, 1997. **47**(10): p. 1145-1151.
21. I.F. Pryme, S. Bardocz, A. Pusztai y S.W. Ewen, *Suppression of growth of tumour cell lines in vitro and tumours in vivo by mistletoe lectins*. Histol Histopathol, 2006. **21**(3): p. 285-299.
22. T. Hajto, K. Hostanska, T. Berki, L. Palinkas, F. Boldizsar y P. Nameth, *Oncopharmacological perspectives of a plant lectin (Viscum album agglutinin-I): Overview of recent results from in vitro experiments and in vivo animal models, and their possible relevance for clinical applications*. Evid.-based Complement Altern. Med., 2005. **2**(1): p. 59-67.
23. T. Ostermann, C. Raak y A. Bussing, *Survival of cancer patients treated with mistletoe extract (Iscador): a systematic literature review*. BMC Cancer, 2009. **9**: p. 451.
24. B.M. Heiny, V. Albrecht y J. Beuth, *Correlation of immune cell activities and beta-endorphin release in breast carcinoma patients treated with galactose-specific lectin standardized mistletoe extract*. Anticancer Res, 1998. **18**(1B): p. 583-586.
25. T. Hajto, K. Hostanska y H.J. Gabius, *Modulatory potency of the beta-galactoside-specific lectin from mistletoe extract (Iscador) on the host defense system in vivo in rabbits and patients*. Cancer Res, 1989. **49**(17): p. 4803-4808.
26. A. de Giorgio y J. Stebbing, *Mistletoe: for cancer or just for Christmas?* The Lancet Oncology, 2013. **14**(13): p. 1264-1265.
27. R. Mikeska, R. Wacker, R. Arni, T.P. Singh, A. Mikhailov, A. Gabdoulkhakov, W. Voelter y C. Betzel, *Mistletoe lectin I in complex with galactose and lactose reveals distinct sugar-binding properties*. Acta Crystallogr Sect F Struct Biol Cryst Commun, 2005. **61**(Pt 1): p. 17-25.
28. K. Drickamer, *Making a fitting choice: common aspects of sugar-binding sites in plant and animal lectins*. Structure, 1997. **5**(4): p. 465-468.
29. J.L. Asensio, A. Arda, F.J. Canada y J. Jimenez-Barbero, *Carbohydrate-aromatic interactions*. Acc Chem Res, 2013. **46**(4): p. 946-954.
30. S. Bharadwaj, H. Kaltner, E.Y. Korchagina, N.V. Bovin, H.J. Gabius y A. Surovia, *Microcalorimetric indications for ligand binding as a function of the protein for galactoside-specific plant and avian lectins*. Biochim Biophys Acta, 1999. **1472**(1-2): p. 191-196.

31. R.T. Lee, H.J. Gabius y Y.C. Lee, *Ligand binding characteristics of the major mistletoe lectin*. The Journal of biological chemistry, 1992. **267**(33): p. 23722-23727.
32. S. Kemper, M.K. Patel, J.C. Errey, B.G. Davis, J.A. Jones y T.D. Claridge, *Group epitope mapping considering relaxation of the ligand (GEM-CRL): including longitudinal relaxation rates in the analysis of saturation transfer difference (STD) experiments*. J Magn Reson, 2010. **203**(1): p. 1-10.
33. V. Jayalakshmi y N. Rama Krishna, *CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities*. J Magn Reson, 2004. **168**(1): p. 36-45.
34. V. Jayalakshmi y N.R. Krishna, *Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes*. J Magn Reson, 2002. **155**(1): p. 106-118.
35. V. Jayalakshmi, T. Biet, T. Peters y N.R. Krishna, *Refinement of the conformation of UDP-galactose bound to galactosyltransferase using the STD NMR intensity-restrained CORCEMA optimization*. J Am Chem Soc, 2004. **126**(28): p. 8610-8611.
36. J. Angulo, P.M. Enriquez-Navas y P.M. Nieto, *Ligand-receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates*. Chemistry, 2010. **16**(26): p. 7803-7812.
37. J. Angulo y P.M. Nieto, *STD-NMR: application to transient interactions between biomolecules-a quantitative approach*. Eur Biophys J, 2011. **40**(12): p. 1357-1369.
38. P.W. Wertz, J.C. Garver y L. Anderson, *Anatomy Of A Complex Mutarotation - Kinetics Of Tautomerization Of Alpha-D-Galactopyranose And Beta-D-Galactopyranose In Water*. J. Am. Chem. Soc., 1981. **103**(13): p. 3916-3922.
39. A.M. Wu, L.K. Chin, H. Franz, U. Pfeller y A. Herp, *Carbohydrate specificity of the receptor sites of mistletoe toxic lectin-I*. Biochim Biophys Acta, 1992. **1117**(2): p. 232-234.
40. M. Mayer y B. Meyer, *Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor*. J Am Chem Soc, 2001. **123**(25): p. 6108-6117.
41. J. Yan, A.D. Kline, H. Mo, M.J. Shapiro y E.R. Zartler, *The effect of relaxation on the epitope mapping by saturation transfer difference NMR*. J Magn Reson, 2003. **163**(2): p. 270-276.
42. N.R. Krishna y V. Jayalakshmi, *Quantitative Analysis of STD-NMR Spectra of Reversibly Forming Ligand-Receptor Complexes*. Top Curr Chem, 2008. **273**: p. 15-54.
43. J. Angulo, I. Diaz, J.J. Reina, G. Tabarani, F. Fieschi, J. Rojo y P.M. Nieto, *Saturation transfer difference (STD) NMR spectroscopy characterization of dual binding mode of a mannose disaccharide to DC-SIGN*. Chembiochem, 2008. **9**(14): p. 2225-2227.
44. R. Krauspenhaar, S. Eschenburg, M. Perbandt, V. Kornilov, N. Konareva, I. Mikailova, S. Stoeva, R. Wacker, T. Maier, T. Singh, A. Mikhailov, W. Voelter y C. Betzel, *Crystal structure of mistletoe lectin I from Viscum album*. Biochem Biophys Res Commun, 1999. **257**(2): p. 418-424.
45. D.S. Goodsell y A.J. Olson, *Automated docking of substrates to proteins by simulated annealing*. Proteins, 1990. **8**(3): p. 195-202.
46. G.M. Morris, D.S. Goodsell, R. Huey y A.J. Olson, *Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4*. J Comput Aided Mol Des, 1996. **10**(4): p. 293-304.
47. A. Ortega, D. Amoros y J. Garcia de la Torre, *Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models*. Biophys J, 2011. **101**(4): p. 892-898.
48. M. Jimenez, S. Andre, H.C. Siebert, H.J. Gabius y D. Solis, *AB-type lectin (toxin/agglutinin) from mistletoe: differences in affinity of the two galactoside-binding*

- Trp/Tyr-sites and regulation of their functionality by monomer/dimer equilibrium.* Glycobiology, 2006. **16**(10): p. 926-937.
49. B. Meyer y T. Peters, *NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors.* Angew Chem Int Ed Engl, 2003. **42**(8): p. 864-890.

ANEXO (Capítulo 4)

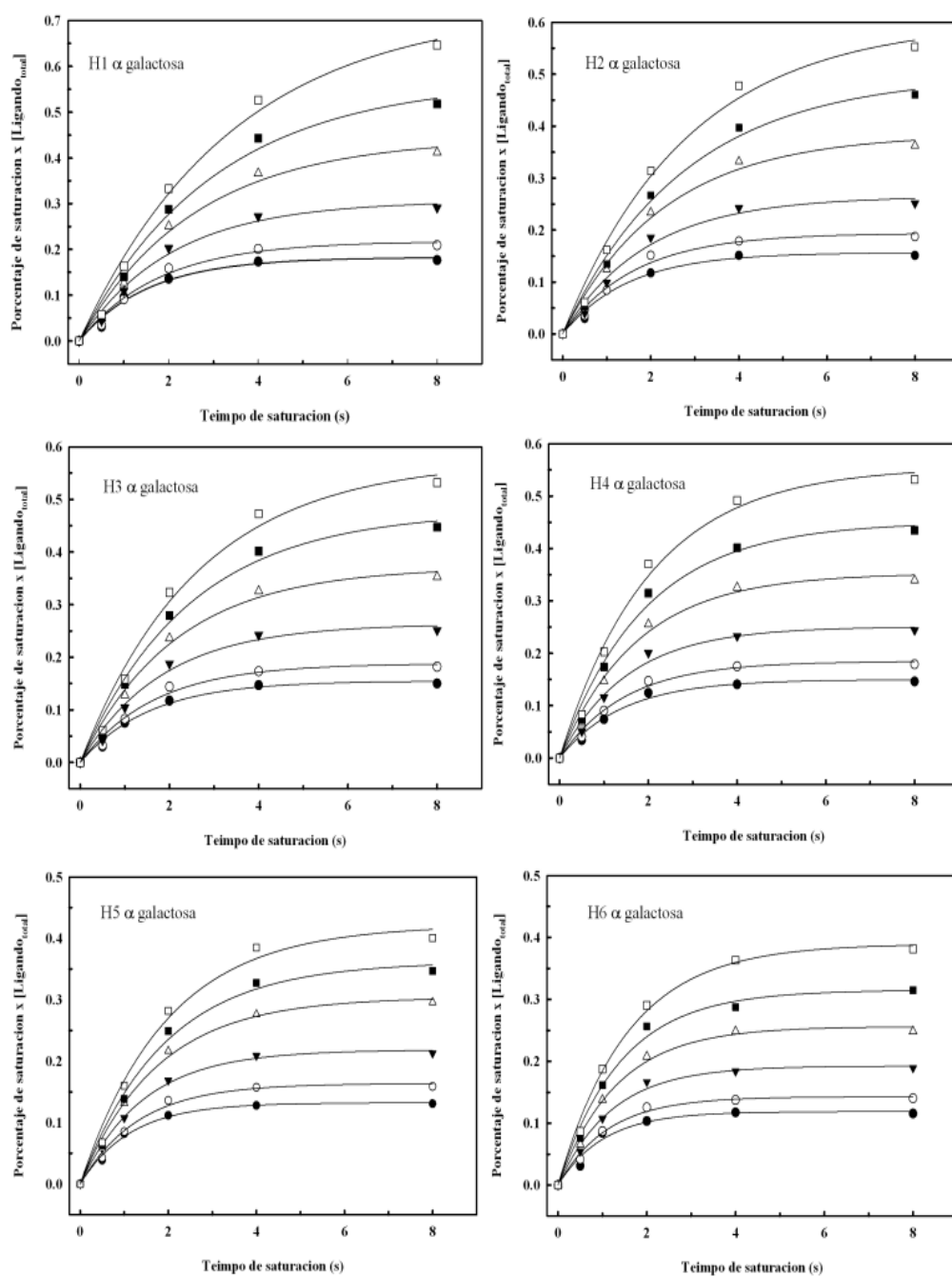


Figura 1 Curvas de crecimiento de saturación en función del tiempo para cada protón del anómero α de la galactosa a diferentes concentraciones de galactosa (\bullet 1; \circ 1,26; \blacktriangledown 2; \triangle 3,6; \blacksquare 6,1; \square 9,8 mM). La concentración del anómero α corresponde a un 33% de la galactosa total.

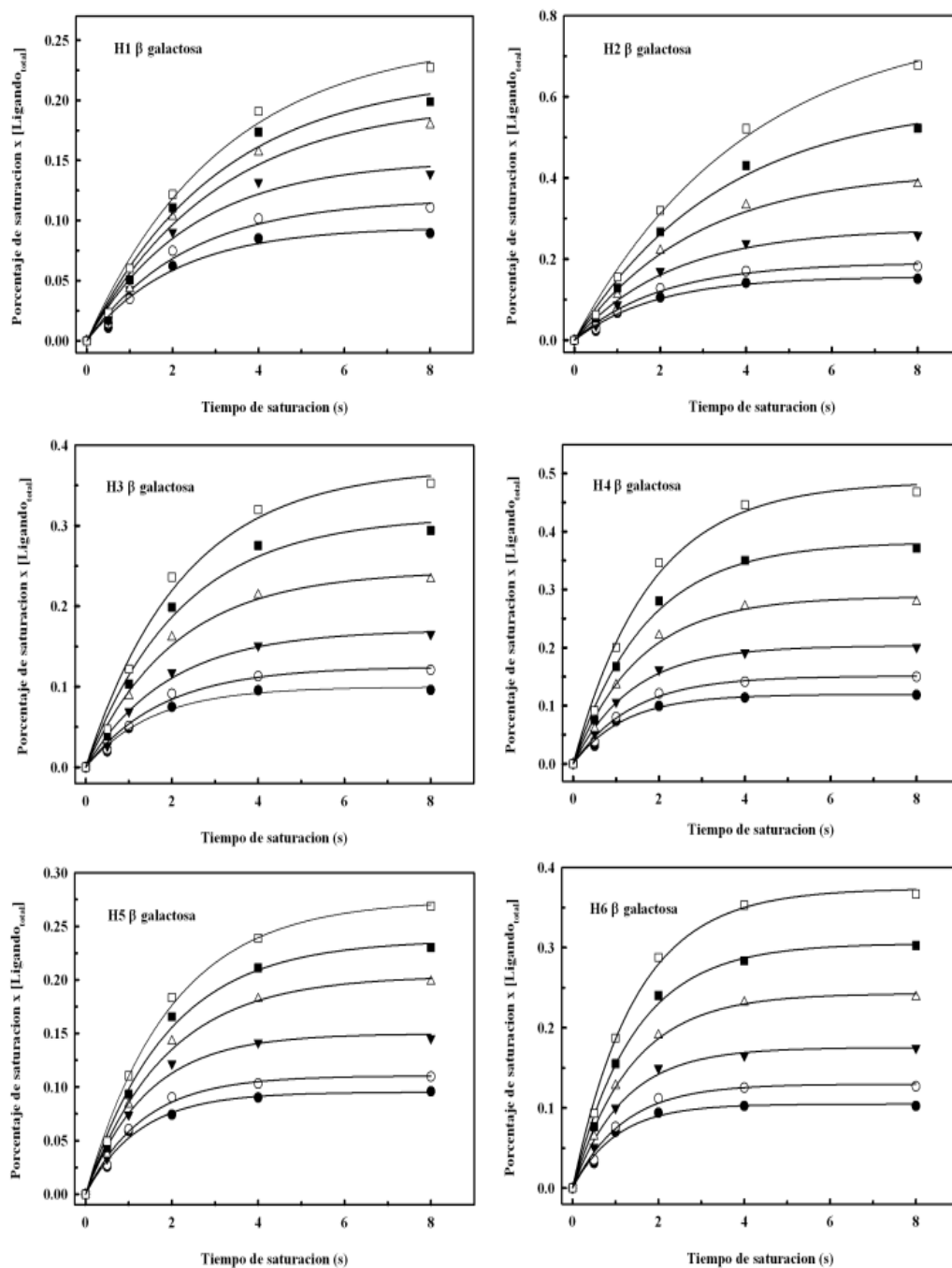


Figura 2 Curvas de crecimiento de saturación en función del tiempo para cada protón del anómero β de la galactosa a diferentes concentraciones de galactosa (\bullet 1; \circ 1,26; \blacktriangledown 2; Δ 3,6; \blacksquare 6,1; \square 9,8 mM). La concentración del anómero β corresponde a un 64% de la galactosa total.

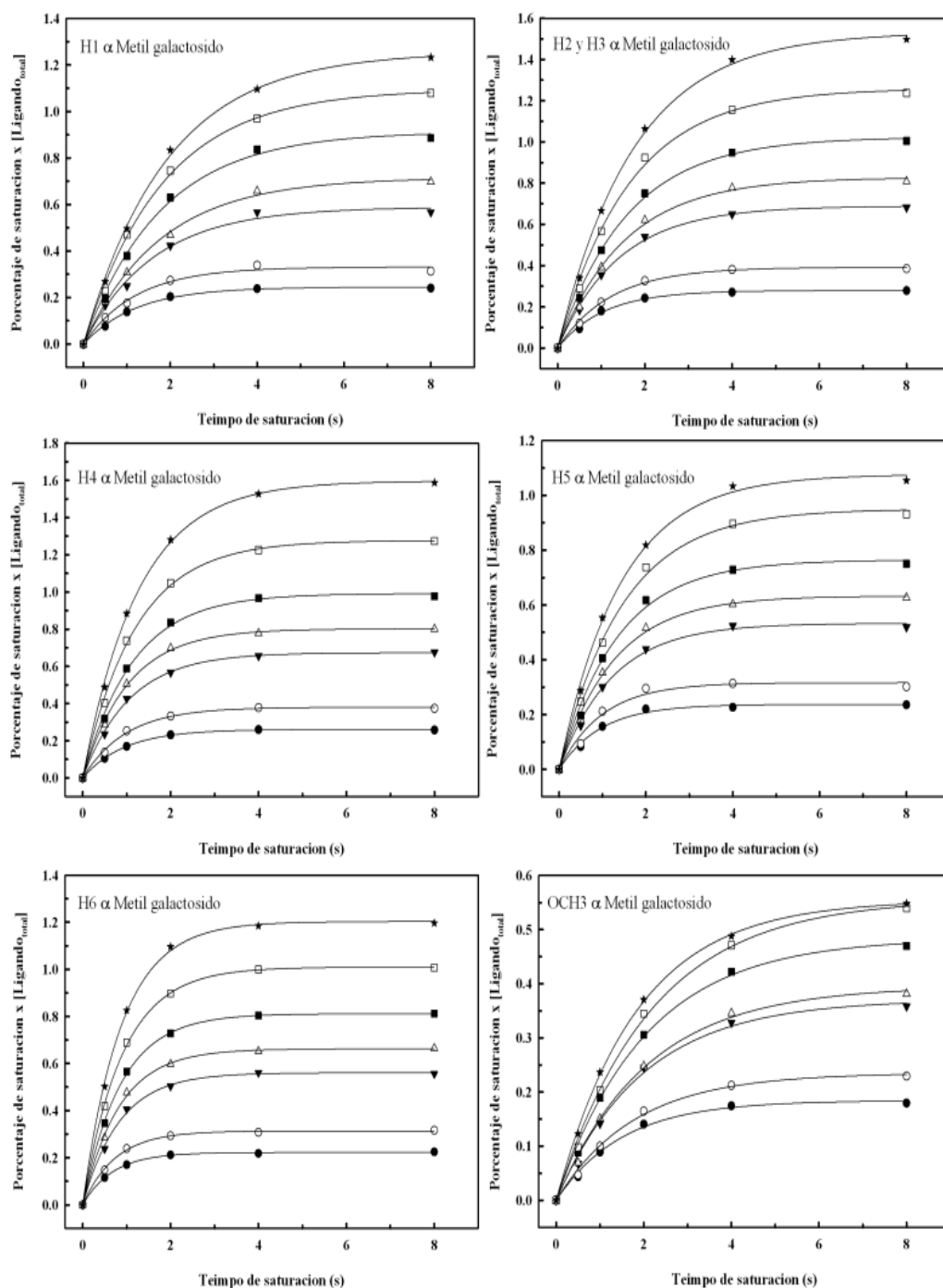


Figura 3 Curvas de crecimiento de saturación en función del tiempo para cada protón del α metil galactósido a diferentes concentraciones (● 0,5; ○ 0,84; ▼ 1,74; △ 2,32; ■ 3; □ 4,4; * 7 mM).

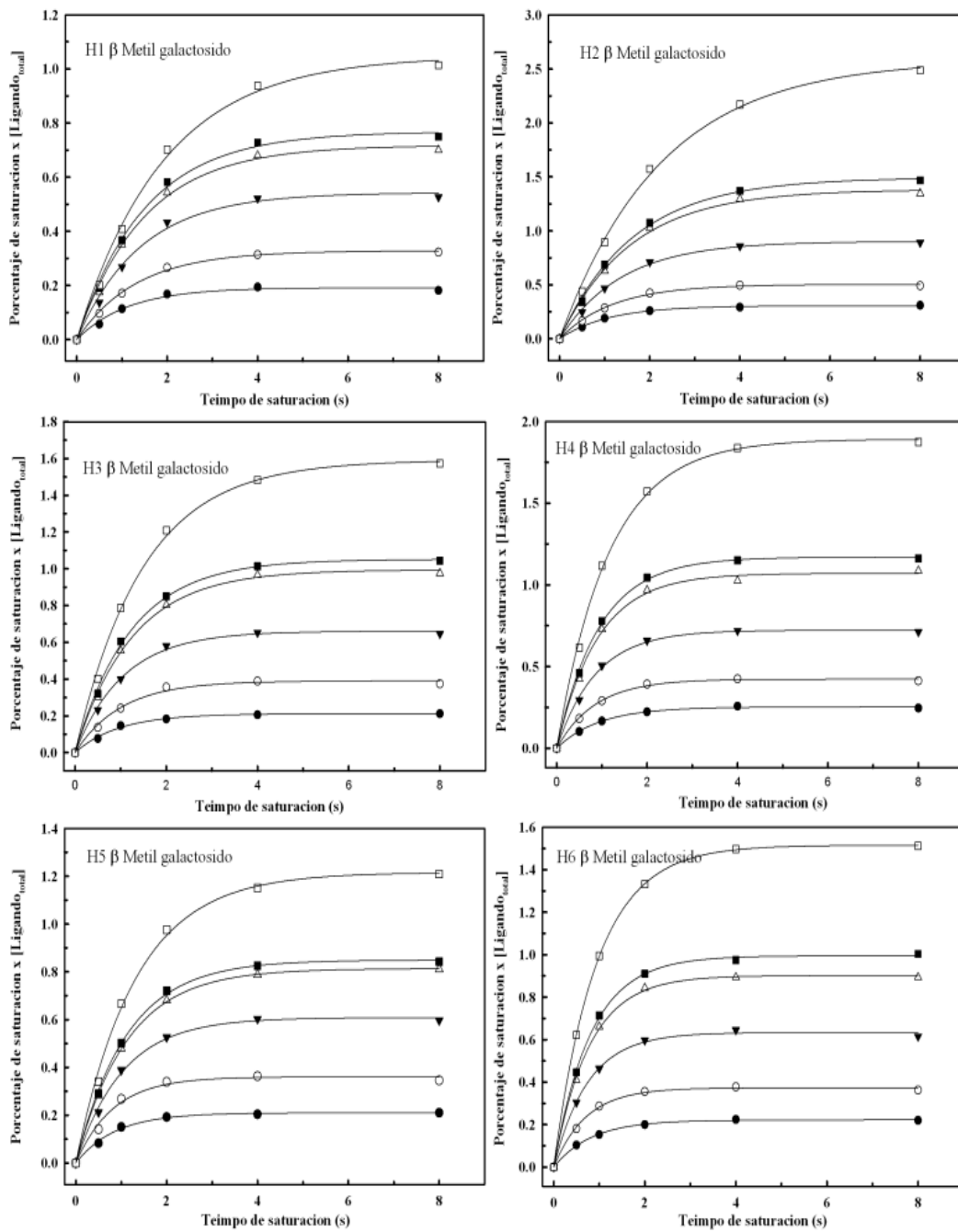


Figura 4 Curvas de crecimiento de saturación en función del tiempo para cada protón del β metil galactósido a diferentes concentraciones (●0,5; ○0,9; ▼1,6; △2,9; ■3,4; □7 mM)

